



CASE STUDY

Open data in the governance of South African higher education

François van Schalkwyk, Michelle Willmers & Laura Czerniewicz

Published in 2014 by OpenUCT
<http://openuct.uct.ac.za/>



This report and its contents are licensed in terms of a **Creative Commons Attribution Share Alike (CC-BY-SA)** license

The funding for this work has been provided through the World Wide Web Foundation 'Exploring the Emerging Impacts of Open Data in Developing Countries' research project, supported by grant 107075 from Canada's International Development Research Centre (web.idrc.ca). Find out more at www.opendataresearch.org/emergingimpacts



Contents

Executive summary	4
Acronyms and abbreviations.....	5
Introduction	6
Structure of this report.....	7
Section 1. Context.....	8
Open data in the national context.....	8
Open data relevant to the governance of South African public universities.....	9
Data and the governance of public universities in South Africa	11
The Centre for Higher Education Transformation (CHET).....	12
Section 2. Methodology.....	14
Research questions	14
Conceptual framework.....	14
Methods.....	16
Selecting university planning departments to interview	17
Selecting higher education studies researchers to interview	18
Data collection instruments	18
Section 3. Findings: Technical platforms and standards	19
CHET	19
HEMIS.....	19
Section 4. Findings: the supply of (open) data	21
Section 5. Findings: data use and impacts	26
Use of CHET open data by university planners.....	26
Use of CHET open data by higher education studies researchers.....	29
Usage based on web statistics	32
User onus and data marketisation.....	33
Conclusion	33
Section 6. Discussion: The higher education governance data ecosystem – data viscosity and role of intermediaries	35
Actors	37
Open data sources.....	37
Policies and legislation.....	38
Context.....	38
The [open] data ecosystem	38
Open data does not exist in isolation	40
Power dynamics.....	40
Keystone species	41
Sustainability.....	41
Capacity	42
Intermediaries and information injustice.....	44
Conclusion	45
Section 7. Summary of findings	47
References.....	49
Appendix 1 The CHET Open Data initiative (http://www.chet.org.za/data).....	52
Appendix 2 Selection matrix for university planning units to include in the sample	54
Appendix 3 The DHET open data website	55

Executive summary

The availability and accessibility of open data has the potential to increase transparency and accountability and, in turn, the potential to improve the governance of universities as public institutions. In addition, it is suggested that open data is likely to increase the quality, efficacy and efficiency of research and analysis of the national higher education system by providing a shared empirical base for critical interrogation and reinterpretation. The Centre for Higher Education Transformation (CHET) has developed an online, open data platform providing institutional-level data on South African higher education. However, other than anecdotal feedback, little is known about how the data is being used. Using CHET as a case study, this project studied the use of the CHET open data initiative by university planners as well as by higher education studies researchers. It did so by considering the supply of and demand for open data as well as the roles of intermediaries in the South African higher education governance ecosystem.

The study found that (i) CHET's open data is being used by university planners and higher education studies researchers, albeit infrequently; (ii) the government's higher education database is a closed and isolated data source in the data ecosystem; (iii) there are concerns at both government and university levels about how data will be used and (mis)interpreted; (iv) open data intermediaries increase the accessibility and utility of data; (v) open data intermediaries provide both supply-side as well as demand-side value; (vi) intermediaries may assume the role of a 'keystone species' in a data ecosystem; (vii) intermediaries have the potential to democratise the impacts and use of open data – intermediaries play an important role in curtailing the 'de-ameliorating' effects of data-driven disciplinary surveillance..

The report concludes as follows: (i) despite poor data provision by government, the public university governance open data ecosystem has evolved because of the presence of intermediaries in the ecosystem; (ii) by providing a richer information context and/or by making the data interoperable, government could improve the uptake of data by new users and intermediaries, as well as by the existing intermediaries; and (iii) increasing the fluidity of government open data could remove uncertainties around both the degree of access provided by intermediaries and the financial sustainability of the open platforms provided by intermediaries.

Acronyms and abbreviations

CHET	Centre for Higher Education Transformation
DHET	Department for Higher Education and Training (South Africa)
DUT	Durban University of Technology
HEMIS	Higher Education Management Information System
NMMU	Nelson Mandela Metropolitan University
ODDC	Emerging Impacts of Open Data in Developing Countries project
UCT	University of Cape Town
UFH	University of Fort Hare
UJ	University of Johannesburg
UP	University of Pretoria
WSU	Walter Sisulu University

Introduction

Higher education has a critical role to play in the development of the nation states of the global South. The creation of new knowledge, innovation, the training of professionals and instilling democratic values are in the hands of the contemporary university (Castells, 2001). However, to be effective, efficient and efficacious in the execution of these obligations universities require access to basic but critical data on, for example, enrolments, graduation (throughput) rates, student composition, student:staff ratios, staff qualifications, research output and impact, sources of income, and more. In particular, *open* data is considered to play a role in improving governance (where governance is considered to be concerned primarily with processes of decision-making and implementation [UN ESCAP, n.d.]) through increasing the transparency of decision-making as well as the accountability of those tasked with implementing processes that serve the public interest.

From a study on the impact of external quality management on universities, Stensaker (2003) points to the fact that the production and sharing of data makes universities more transparent, opens up the 'black box' of higher education and, in so doing, improves decision-making. This finding is supported by studies that have found that external quality management may have unanticipated impacts on both organisational and management issues. Greater centralisation of procedures and organisational decision-making is one trend noted by these studies. A more autonomous role for university management, including affording managers greater responsibility for taking action was another trend noted. As Stensaker (2002:10) concludes: '[G]reater autonomy and increased centralisation of higher education institutions have, in general, given the institutional leadership more visibility and power to play a more important role in policy implementation and organisational change processes as an interest negotiator, a policy translator and as a creator of meaning.'

The increase in managerialism in higher education on a global scale is well-documented (Amaral et al., 2002; Bentley et al., 2006; Brennan, 2008; Reed et al., 2002). However, managerialism without informed decision-making has the very real potential to foster weak and fragmented institutions prone to corruption and/or the inappropriate allocation of resources. This potentially destructive combination is among the reasons for 5 of the 23 public universities in South Africa being under administration at the time of writing.¹ South African university councils need accurate and informative data on the state of their institutions in order to shift the debate from one that is driven by ideology and self-interest to one that is empirically based and in the interest of the performance of the institution.²

This report shares the findings from research that sought to examine the supply, (re)use and possible impact of open data on the governance of South African public universities. To do so, interviews were conducted with university planners as central actors in the (re)use of open data in university governance. Acknowledging their role in the research-policy nexus, the research also considered the use of open data by higher education studies researchers. The Centre for Higher Education Transformation's (CHET)³ open data platform was used as a case to examine the dynamics of data supply, (re)use and the role of intermediaries. The research formed part of a larger 17-project research project, the Exploring the Emerging Impacts of Open Data in Developing Countries (ODDC), funded by the Canadian International Development Research Centre (through grant 107075) and managed by the World Wide Web Foundation.

¹ See Sunday Independent, 29 July 2012, 'Poor leadership cripples tertiary institutions'.

² This need is confirmed by the number of requests received by the Centre for Higher Education Transformation in 2013 from four South African public universities (and three African universities) to present to council (and management) a set of institutional-level indicators following the completion of its research project on national and cross-national performance indicators.

³ www.chet.org.za

Structure of this report

This report is structured in seven main sections.

The first section begins by providing an overview of the open data context in South Africa, and the context of university governance in South Africa. This explains the relevance of open data to university governance, and briefly describes the history of the CHET initiative which formed the core focal point for this case study.

The second section provides an overview of the research questions, conceptual framework and research methodology, justifying the choices of interview sites and methods in light of the governance and open data contexts of the study.

The next three sections present empirical findings from the core case study of the CHET initiative, starting with a brief overview of technical platforms and standards is provided. This component of the conceptual framework, although no less important, was not studied in-depth in this research project. The second findings section situates CHET within a wider landscape of open data provision for higher education governance, before the final case study finding section explores in depth qualitative insights from users or potential users of the CHET open data. Through this layout of the findings, supply is discussed separately (although not without reference to demand-side dynamics) with this separation of open data supply in the analysis allowing for a simpler and more defined entry point into a complex domain of (open) data supply, demand and (re)use.

The sixth section steps back from this case-study narrative, to revisit the CHET case study in the context of a much wider ecosystem of higher education governance data. This final section of the report attempts to capture some the complexity in the supply, (re)use and impact of open data that was limited by the deliberately bounded supply-side case study in the previous sections. Using the heuristic of the ecosystem, a more nuanced and holistic picture of the roles of intermediaries within the ecosystem is presented. This section draws on additional empirical evidence, and unpacks a number of key theoretical and practical points for consideration when approaching open data projects.

The report concludes with a final section providing a collation of the key findings and a corresponding set of recommendations.

Section 1. Context

Open data in the national context

Internationally, the open data movement has gained significant momentum in recent years. In the case of national government data, South Africa is a member state of the Open Government Partnership. The South African government's action plan submitted to the Open Government Partnership focuses on inclusion and consultation as key drivers of accountability, transparency and service delivery but makes no specific reference to opening open up or linking government-owned data in support of these objectives. Only one of the eight implementation commitments made the South African government – the commitment to investigate the feasibility of an open environmental management information portal – could be described as implicitly relying on open data and interoperability for its success. Government's commitment to e-government is expressed in its "e-Government Strategic Framework: Accelerating Service Delivery 2014" and the associated implementation framework "Minimum Interoperability Standards for Government". However, both the self-evaluation reports and external reviews show that what is transpiring by way of implementation is way below expectations.

At the departmental level, the Department for Higher Education and Training (DHET) is engaged in a large-scale project which aims to integrate data from multiple sectors and sources to develop a responsive, data-driven labour market intelligence system.⁴ The Department's commitment to a single, integrated information system is set out in its most recent policy document, the 2013 White Paper White Paper for Post-school Education and Training: Building an Expanded, Effective and Integrated Post-school System (Department of Higher Education and Training, 2013). There is no evidence to suggest that the data collected will be in the public domain. Nor is there much evidence of open data initiatives in other government departments or agencies.⁵ For example, the Open Data for Africa Portal lists 26 datasets on providing open data on South Africa.⁶ Only one South African government department contributes data, Statistics South Africa (8 datasets), making this department the exception that proves the rule.

Statistics South Africa is also the government agency responsible for the quality of official national statistics, including statistics generated from data provided from other government departments. According to the Statistics South Africa website, "[a]n increase in the demand by government departments and other data producers was due to the demand for statistics that can be trusted. SASQAF [South African Statistical Quality Assessment Framework] was used as an enabler for self-assessment by data producers; mainly organs of state, review performed by a Data Quality Assessment Team (DQAT) in context of the National Statistics System (NSS) for certifying statistics as official as stipulated in the Statistics Act (Act No.6 of 1999). The Statistics Act mandates the Statistician-General to formulate quality criteria and establish standards, classifications and procedures for statistics produced by all organs of state and other agencies that produce statistics and to designate as official, statistics or class of statistics produced by any organ of state."⁷ The South African Statistical Quality Assessment Framework includes criteria many of which are promoted by the open data movement: data relevance, quality, timeliness, accessibility, interpretability, compliance and coherence. Specifically, in the section on accessibility, the Framework includes requirements on public dissemination of data and the existence of a dissemination policy, on equality of access, and on the supply of multiple data formats. However,

⁴ See the Labour Market Intelligence Project (LMIP) website www.lmip.org.za.

⁵ See http://www.dailymaverick.co.za/opinionista/2013-04-12-public-data-in-south-africa-time-to-claim-whats-ours/#.U9iUy_mSys0

⁶ <http://southafrica.opendataforafrica.org/data#menu=source>

⁷ http://beta2.statssa.gov.za/?page_id=371

SASQAF provides no explicit requirements in terms of unfettered data reusability nor on any requirements for data to be interoperable.

At the regional (provincial) and city levels, there is little activity – none of South Africa’s major metropolitan cities have an open data portal, although in some cases metros have initiated processes to explore the possibility of opening up their data, and there is evidence of open data policies being drafted at city level.⁸ Other policies relating to open standards and open software were found to exist at various government levels, but they appeared to be having little impact. No specific policies on open data were identified.

The drive for government and publicly-funded institutions to release data on their operations, research and expenditure has, however, been fuelled by international protocols and the drive for increased accountability. In the higher education and research environment, a parallel movement around making research data openly available has gained increasing support in light of transformative scholarly communication activity that now encourages the curation and sharing of all products generated in the knowledge production cycle for the purposes of reproducibility, efficiency and transparency. While the drivers for these two sectors of open data activity align in many respects, the communities they engage are often quite different.

The open data movement in South Africa is relatively nascent, both with respect to sharing government and research data. This trend does however show signs of change as international scholarly communication practice and content-sharing mandates from the global North filter into local practice (particularly affecting those who collaborate internationally or whose work is funded by mandating agencies). The controversial Protection of State Information Bill, which is popularly referred to as the ‘Secrecy Bill’, has been formulated to stifle press freedom, and also acted as a significant driver in mobilising data activists who are “turning up the volume on their demands for government and publicly funded institutions to release data”.⁹ At the same time, there are efforts afoot to train data journalists, and non-governmental organisations such as Code4SA are providing support to civil society in opening data and in advocating for a more comprehensive corpus of government data. But it is still early days in South Africa. As a press release from as recent as 10 July 2014, following a national meeting of stakeholders, reads: “Over 90 of South Africa’s open data experts from government, civic organisations and private enterprise met at the #OpenDataNow Unconference in Cape Town last week **to take the first steps** in creating an open data society. Delegates promised, amongst other commitments, to build a government API catalogue and create a South African data visualization community for open data experiments” (emphasis added).¹⁰

Open data relevant to the governance of South African public universities

This overview provides a glimpse of the South African higher education open data activity. It provides a sense of how mature or nascent the national open data environment as it pertains to university governance is, how it fares against international definitions of openness and points to criteria of openness not met across the data sources identified.

⁸ For example, see the City of Cape Town’s Draft Open Data Policy <http://www.capetown.gov.za/en/publicparticipation/pages/pendatadraftpolicy.aspx>, and initiatives in Gauteng-region cities around opening spatial data https://ujdigispace.uj.ac.za/bitstream/handle/10210/8226/Wray%20%26%20Olst%20_2011.pdf?sequence=1

⁹ See <http://www.ip-watch.org/2013/05/08/a-battle-for-open-public-data-in-south-africa/>

¹⁰ <http://www.code4sa.org/2014/08/10/pendata-unconference-press-release.html>

Desktop research was done in order to identify open data initiatives in the South African higher education context. The investigation began with the data already known and analysed as part of this study, that is, the Centre for Higher Education Transformation's open data portal and the open data provided by the South African Department for Higher Education and Training. Based on these core data sources the websites of other government departments, higher education bodies and known data repositories in South Africa were analysed. International organisations, known to be providers and sources of data, were also included in the investigation. Although the focus was to consider open data initiatives in the South African context, it was noted that international data providers extracted data from South African data sources; therefore including international organisations such as the World Bank and the United Nations could be useful in identifying South African data suppliers.

The first step was to cast a broad net in order to identify available South African open data sets; the next step entailed scrutinising the data and associated platforms in order to gauge whether they could be considered 'authentically' open in line with international protocols and definitions.¹¹ The conditions for openness of two different interest groups were applied: Open Government Data Principles which were developed by a working group of open government advocates and a second set of principles set out by the Open Data in Developing Countries (ODDC) Initiative.

The subsample of South African higher education data can be summarised as containing 25 datasets, including the Higher Education Management Information System (HEMIS) data tables, the AISA geo-portal and the Department of Science and Technologies R&D Survey Results, as well as individual studies and surveys such as the CHEC Graduate Pathways Study.¹² Seventeen data providers, including DHET, CHET, Statistics SA, South African Data Archive (SADA) and DataFirst were identified.

Each dataset in the analysis was assessed according to each of these criteria and given a score out of 10 (for the ODDC criteria) or 8 (for the Open Government Data criteria). A decision was made by the research team that a dataset would be considered "open" if it met at least 80% of the criteria using each of the evaluatory frameworks. The decision to settle for a percentage score in order to award "open" status was taken after initial evaluation revealed that none of the datasets would be considered open if a score of 100% was required. In light of the fact that open data practice was still relatively nascent in the South African context and was being enacted as a result of a wide range of drivers and parallel incentives, it was deemed important to acknowledge openness of data in whatever context it manifest – in part to identify the areas of activity where open data practice is developing, but also to support the practice wherever possible rather than judging explicitly against international checklists. The items of the two evaluatory frameworks which are not currently being met provide some indication of areas where additional work is required in growing and formalising practice.

In terms of the ODDC ten-point framework, which places an emphasis on governance in a developing country context, many datasets were particularly weak in the following areas: machine readability, availability of data in bulk, licensing being unclear and/or not openly licensed, and no linked data URIs provided. The failure to meet the linked data URIs criterion was found across the board. Only one dataset managed to meet 80% of the ODDC criteria – the CHET higher education data. The CHET data met nine of the ten criteria, failing only on the linked data URIs criterion.

When using the Open Government Data principles, which focuses more on government transparency and accountability, the following criteria were not met by the datasets: machine readability, unclear and/or restricted licensing of data, and non-discriminatory access to data. Seven datasets were found to be open

¹¹ This exercise was similar but more extensive version of the study undertaken by Van Schalkwyk (2013) in examining the flow of open data using the case of public university governance in South Africa in that included a much larger sample of datasets.

¹² For full breakdown of sources, resources and providers, see source data available at: <https://zenodo.org/record/10006#.U416cC8rzYY>

using the Open Government principles. All seven of these come from government departments or government agencies, bodies and councils.

This exploratory exercise suggests that the open data movement in South African higher education data and research currently holds fledgling status and requires additional efforts to fall in line with international protocols and open standards. Locally CHET is a leader in sharing data and information in the higher education governance context, but it still requires additional time and development to meet the principles for transparency and accountability. Machine readability and licensing of data sources and resources are also significant challenges facing South African higher education open data movement – and indeed the local scholarly communication enterprise as a whole. These criteria form a part of both the ODDC and Open Government Data frameworks and are thus central to the principle of openness, particularly because they facilitate the ability to use, (re)use and redistribute data efficiently (principles central to the [Open Definition](#)).

Despite growing activity and signs of increased uptake, preliminary evidence suggests that open data activity in South Africa is still relatively haphazard. What is clear, however, is that there is a paucity of research in this area.

Data and the governance of public universities in South Africa

South African higher education is regulated by the provisions of the Higher Education Act of 1997. The Department of Higher Education and Training (DHET) is the government department responsible for the public higher education sector. Universities are largely autonomous and government steers the system by setting goals at the system and institutional levels, and by monitoring the performance of the system and of individual institutions against these goals (Bunting et al., 2010). The primary state steering lever is government funding in the form of annual block grants and earmarked funds for designated projects (Pillay, 2010). In order to monitor performance, universities are required to submit data for inclusion in the Higher Education Management Information System (HEMIS), which is managed by DHET. Universities are required to capture and submit HEMIS data on students, on staff and on building space data.¹³ HEMIS forms the basis for annual state funding allocations as well as system-level policy/steering decisions.¹⁴ However, while publicly funded and technically in the public domain,¹⁵ the data is not easily located nor is it available publicly in a usable format.¹⁶ Furthermore, it appears that at national level, DHET lacks the internal capacity to curate and share the data in a manner that ensures optimal discoverability, (re)use and impact.

At the level of the university, governance structures are largely homogenous across the system of 23 public universities. The Act outlines the functions of university governance structures and asserts the supremacy of the council as the final authority in university-level governance (Ncayiyana & Hayward, 1999). Council is advised by university executives who head the various organisational units. Typically, it is the function of the institutional planning unit (or its equivalent) to provide strategic support services to the university executive, including the provision of management information that is both relevant and timely for strategic decision-making. While university planners have access to their own captured data

¹³ Universities are also required to submit finance and research output data but this data is not stored in the HEMIS database. Finance and research data are submitted in the form of annual reports.

¹⁴ For example, extensive use is being made of the HEMIS database by a committee that has been convened by DHET to review the current higher education funding formula (Interview).

¹⁵ Student and staff data of a personal nature is treated as confidential and is therefore not available in the public domain.

¹⁶ See <http://www.dhet.gov.za/Structure/Universities/ManagementandInformationSystems/tabid/419/Default.aspx> for the data currently available on the DHET website. The data has only recently been made available on the website, is hard to locate on the site and consists of large Excel tables typically annotated with indecipherable category codes and abbreviations.

and to report-writing tools within university management information systems, comparative data is not available.

In the case of those conducting research on South African higher education, the HEMIS database appears to be difficult to access without existing relationships in place and the open data tables appear to be difficult to interpret without prior knowledge of how the data is coded and structured by DHET.

In other words, while there is a rich publicly-funded dataset on South African higher education, the data remains largely inaccessible or unusable to universities and researchers in higher education studies. This raises questions around the value and usefulness of the DHET HEMIS data in its current form to the governance of higher education in South Africa. While opening up datasets is an important first step, 'data dumping' may hold little value for academics and/or practitioners.

The Centre for Higher Education Transformation (CHET)

In 1999, the Centre for Higher Education Transformation (CHET) initiated a project on performance indicators in South African higher education. The project arose from the question: 'Is the South African higher education system transforming?' By 2000, the concept of 'transformation' had become so ideologised, it argued, that the concept no longer had any empirical meaning. It maintained that only a combination of empirical indicators and theoretical reflection would initiate constructive dialogue between stakeholders on the transformation of the South African higher education system.

The first publication from the performance indicators project was published by CHET in 2000, *Higher Education Transformation: Assessing Performance* (Cloete & Bunting, 2000).¹⁷ The report attempted to monitor the progress of transformation, defined mainly in terms of the South African government's 1997 White Paper on Higher Education. Following ongoing work on performance indicators, CHET published in 2004 *Developing Performance Indicators for Higher Education* (Bunting & Cloete, 2004)¹⁸ which explored approaches to the assessment of performance, and by implication, performance indicators from 1999 to 2004 in South Africa. In 2009, based on feedback from universities and the refinement of the indicators proposed in the 2004 publication, CHET published *Performance Indicators: The South African Higher Education System 2000–2008* (Bunting et al., 2010).¹⁹ The book included profiles of South Africa's 23 public universities as well as peer comparisons presented as part of CHET's ongoing work on the measuring of performance within the South African government's steering model of higher education governance. For the first time, the data were also made available on the CHET website via a custom-built platform allowing users to select one of 20 predefined indicators, to select up to four universities to compare, and to generate graphs and downloadable Excel tables (see <http://www.chet.org.za/data/>). Government-set targets were also included in the database and users provided with a glossary of terms relevant only to the graph generated in order to provide definitional clarity. The intention was for the performance profiles, in conjunction with the online open access data, to assist university planners and councils to make assessments that would contribute to evidenced-based management and governance.

In all of its work on higher education performance indicators, CHET drew on data from HEMIS. CHET's role could therefore be described as that of an intermediary as it brokered data between the primary centralised data source (DHET's HEMIS database) and end-users (university planners, higher education studies researchers and others).

¹⁷ <http://www.chet.org.za/books/higher-education-transformation>

¹⁸ <http://www.chet.org.za/books/developing-performance-indicators-higher-education>

¹⁹ <http://www.chet.org.za/books/performance-indicators-south-african-higher-education-2000%E2%80%932008>

CHET's work on performance indicators was intended not only to contribute to the then Department of Education's steering mechanisms (involving planning and funding) aimed at transforming the public higher education system, but also to contribute to the governance capacity of South African universities at the planning and council levels. For if the process of steering by planning and funding is to function effectively, it is crucial that universities understand what is implied by the targets set by government, understand how performance is assessed by the Department, and have both the frameworks and the data available to make empirically-based decisions.

The CHET open data portal exists as an interesting case in point, in that it constitutes a data source pertaining to both research and public administration. It is therefore potentially useful to consider it in terms of the position it inhabits at the intersection between scholarship and governance; appealing to different audiences (managers and scholars) who require the data for quite different purposes (administration and higher education studies). This is particularly interesting in the Southern African higher education context as it feeds into a new strategic drive by many higher education institutions in the region to develop a more data-driven (business intelligence) approach to administration and research management.

Section 2. Methodology

Research questions

Based on the context outlined above, the questions this study proposed to answer are:

1. What is the value of the CHET open data initiative to university planners as well as higher education studies researchers?
2. What is the degree of uptake among selected South African universities and higher education studies researchers?
3. What are the limitations of the CHET open data initiative in its current form?

In addition, this study set out to examine the role of intermediaries as the CHET case was likely to provide unique insights into the complex role between suppliers, intermediaries and consumers of open data on the governance of South African higher education. Universities as both suppliers and consumers, and government as both the primary, centralised public source and consumer of the data, illustrates some of this complexity. Given that the data has always been theoretically open and that universities have always effectively had access to their own data, the motivations of the intermediary in this data flow scenario and extent to which the open data is, to use Yu and Robinson's (2012) terms, 'adaptable' or 'inert', may also provide useful insights.

Conceptual framework

Helbig et. al (2012) argue that research into open data should investigate the context and dynamics within which open data is embedded, including all the "stakeholders, data sources, data resources, information flows, and governance relationships involved in the provision and use of government-held and non-governmental data sources". The research framework of this study follows this recommendation.

It also draws heavily on the conceptual framework developed the Exploring the Emerging Impacts of Data Project of the Web Foundation (Davies et al., 2013) which, following Helbig (2012), is premised on the recognition that the emerging impacts of open data are realised within specific circumstances and contexts. The ODDC framework emphasises an understanding of decision-making in-context, and on exploring the mechanisms by which open data may be a driver for change in distinct governance settings.

In order to extrapolate insights and findings that are contextually bound to more generalisable statements about open data, the ODDC's framework draws attention to three terms: open data, governance and emerging outcomes. The approach recognises that there are many subject areas for open data, and that it is important to understand the subject, structure and status of a dataset in constructing an account of open data impacts. It further recognises that governance issues exist and are addressed at many levels -- political, economic and social. The ODDC framework highlights a range of 'emerging outcome' mechanisms through which open data may deliver impact. The ODDC framework is operationalised by setting out a series of components to include in open data case studies. It is argued by the ODDC that any case study might involve different kinds of data, governance issues and emerging outcomes, and will be responding to local policy and practice questions as well as cross-cutting research issues. For this reason, ODDC highlights six areas that the case studies should address, laying the foundation for a holistic comparative cross-case analysis:

1. The context for open data – including the political, organisational, legal, technical, social and economic contexts.

2. The supply of open data – including data availability, legal frameworks for data, data licenses, and stakeholders involved in providing data.
3. Technical platforms and standards – including data formats and data standards use, as well as data catalogues, APIs or analysis tools provided by an open data initiative.
4. The context of the specific governance setting – including a description and history of the issues in focus, details of key stakeholders, and analysis of how data plays a potential role in this setting.
5. Intermediaries and the flow of data – the means by which data is made accessible in the governance setting.
6. Actions and impacts – the experiences of those seeking to use data, and providing evidence of intended or unintended consequences.

Figure 1 illustrates these six core case components. The context in which data is being supplied as well as the governance context have been introduced in the opening sections of this report. The emphasis of the research and of the analysis presented is on supply, use/impact and the role of intermediaries in the case of the CHET open data initiative. Less attention is paid to technical platforms and standards.

In its simplest form, governance is understood to be concerned with the processes of decision-making and implementation (UN ESCAP, n.d.). For the purposes of this study, public university governance data is presumed to include data on the following:

1. students (enrolment data, graduation data and demographic data such as gender and race),
2. staff (number, type, level and qualifications of staff as well as demographic data),
3. knowledge production (number, type, frequency of publications), teaching and engagement with external stakeholders (type, number of projects, outputs),
4. curriculum (number and types of courses and qualifications offered),
5. space (infrastructure, facilities, equipment),
6. finance,
7. responsiveness to labour market needs, and
8. international university rankings.²⁰

These data represent, in our view, the type of data that universities would use to inform decision-making and implementation at institutional-level. The list may not be exhaustive but it includes all of the data types used by government to steer the public university system, as well two additional data types that are currently topical and often referred to by university planners and leadership.²¹

We recognise that defining “open” and “data” represents a significant challenge in an era where internet-enabled information exchange has introduced varying legal and social interpretations of openness, and data manifests in myriad forms. For the purposes of this investigation, the focus will be on data contained in databases in both “raw” and “processed” or “shaped” forms. Our definition of open data is that formulated by the Open Knowledge Foundation: ‘A piece of data or content is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike.’²² In line with this definition, the concept of “open data” requires that the dataset be available (online), without cost to the user and with no technical or legal restriction to prevent its re-use. For open

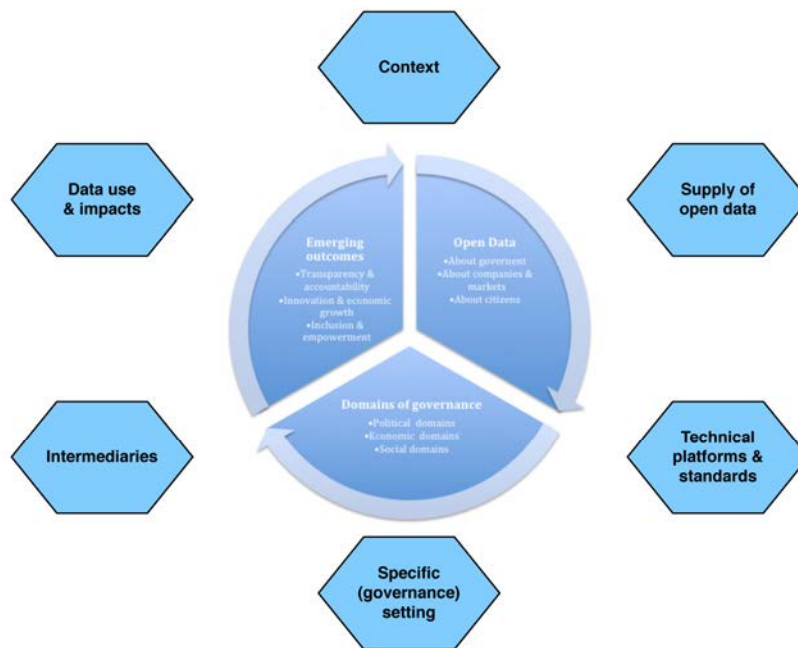
²⁰ The preoccupation of university leadership with university rankings, and their influence on strategic decision-making, has emerged strongly in interviews conducted with planners at South African universities.

²¹ The two “additional” data types are labour market data and global rankings of universities data. In the case of labour market data, the DHET has commissioned the Human Sciences Research Council to develop a labour market intelligence system to allow government, industry and universities to improve the responsiveness of universities (and colleges) to the needs of the labour market. In interviews conducted with university planners, they consistently mention requests from university leadership for data on how their institution is fairing according to the criteria of one of the global university rankings.

²² <http://opendefinition.org/>

government data, the Open Government Data Definition's 8 Principles of Open Government Data (also referred to as the Sebastopol Principles)²³ further require that data be made available in a timely manner and be primary (as collected at the source).²⁴

Figure 1: Exploring the Emerging Impacts of Data Project Conceptual Framework



Methods

This research project on the use and impact of open data in the governance of South African universities adopted a case study approach. According to Clark (2008), case study research (specifically research which is concerned with change in universities) offers four distinct advantages over other approaches, principal among which is that it “commits to local context”; it retains the interplay between the stabilizing and transformative forces of change by retaining the focus of the research on the local context. Helbig et al. (2012) argue that research into open data should investigate the context and dynamics in which open data are situated. In order to develop an understanding of decision-making in different settings and to explore the different mechanisms by which open data may be seen to affect change in distinct governance settings, this project has opted for a case study approach.

Data on the supply and use of open data in South African public university governance was collected from the following, each key in the flow of data from supply to end-use:

1. a representative from the Department of Education and Training (DHET) responsible for the HEMIS database;

²³ <http://www.opengovdata.org/home/8principles>

²⁴ See Van Schalkwyk (2013) for a discussion on how the criteria for openness are contingent on the authoring group's theory of change, i.e. their conception of how opening up data will impact on an existing practice. Interestingly, Manyika et al. (2013) limit their analysis of the economic impact of open data to only four of the eight OGP principles, ostensibly because the criteria they have jettisoned (including timeliness and granularity) are not in their view critical conditions for open data to make an economic impact.

2. the Director of CHET, administrative staff and consultants who have been involved in the performance indicator project since its inception; and
3. users of the CHET open data platform, including
 - a. staff members of university planning departments,
 - b. researchers in higher education studies, and
 - c. other users (media, consulting companies, public, etc.)

The project also had access to analytics data on the CHET website. This data was analysed as an additional indicator on the use of the CHET open data platform.

Selecting university planning departments to interview

One of this research project's primary objectives was to establish whether open data is in fact impacting on the governance of public universities in any way. In other words, are university councils (which the South African Higher Education Act of 1997 asserts as the supreme authority in the governance of universities) using open data in their decision-making and strategic planning?

In order to establish levels of awareness and use, key staff in the institutional planning units at 7 of the 23 South African public universities were interviewed. An initial challenge was the selection of the seven universities to make up the sample. If universities in South Africa were known to have different governance structures across the system, then one approach would have been to select a representative sample across the different modes of governance, and this would have provided the opportunity to examine whether different forms of public university governance are more or less receptive to the use of open data in their decision-making. However, governance structures in South African public universities are fairly homogenous. A different set of selection criteria had to be devised that would nevertheless produce a representative sample of universities. While the sample could not be differentiated by mode of governance, other non-governance differentiators were identified: size (small/medium/large); type (research/technical/comprehensive/distance); location (rural/urban); and multi or single campus. Appendix 2 provides the selection matrix incorporating these criteria.

One 'governance criteria' was, however, identified to inform the selection process. Five South African public universities are currently under administration, four for what can be considered to be for issues around poor governance. We therefore determined to include at least one of these institutions in the sample in order to assess whether we could identify any marked difference in the use of open data for governance at an institution where governance is considered to be problematic.

Based on the above selection criteria, the sample was made up of the following seven public universities:

1. Durban University of Technology (DUT)
2. Nelson Mandela Metropolitan University (NMMU)
3. University of Cape Town (UCT)
4. University of Fort Hare (UFH)
5. University of Johannesburg (UJ)
6. University of Pretoria (UP)
7. Walter Sisulu University (WSU)

Although the project set out to interview university planning departments, it was found that, in practice, the universities interviewed had different structures in place to manage the collection and supply of data.

In the case of DUT, the research director was interviewed as there is no institutional planning unit at DUT and the research director appeared to deal with data requests from DUT management; data collection

and submission of data to DHET is the responsibility of a dedicated HEMIS unit. At UP a dedicated unit for data management, BIRAP, exists within the institutional planning department: “The Bureau of Institutional Research and Planning (BIRAP) assists the Executive and senior management of the University in this endeavor by providing a professional and strategic support service. BIRAP focuses on rendering a specialist service by providing management information to the Executive that is both relevant and timely for the making of strategic decisions.” A similar arrangement was found to be in place at NMMU. At UJ, UCT and UFH, data collection and supply was found to be the responsibility of the institutional planning department as initially anticipated by this project. In the case of WSU – the only university in the sample under administration – a consultant seconded to the university by the government-appointed administrator to oversee the institutional planning (including data collection, analysis and submission) was interviewed.

Selecting higher education studies researchers to interview

Researchers with an interest in South African higher education studies were identified by scanning for articles in prominent higher education journals and other research outputs that clearly used HEMIS data in its research. The project also identified researchers from emailed data requests submitted via the CHET open data website. In all, the project collected data from six higher education researchers, and from six researchers who submitted data requests via the CHET open data website.

Data collection instruments

Data was collected primarily via semi-structured interviews from the representative samples identified above. While time-consuming and therefore limiting in terms of the sample size, semi-structured interviews were preferred as they provided the flexibility required to explore areas unanticipated in the research design while at the same time providing the structure required to collect comprehensive and comparable data. This flexibility was important given that research is being conducted on *emerging* impacts in a new and under-researched area.

Two questionnaires were developed to guide the interview process: one for the university planners and one for the higher education studies researchers. The questionnaires were designed so as to elicit information from the respondents that would inform the research questions, i.e. the uptake of the CHET open data platform, value of open data, and the limitations of the CHET open data.

Data collection from those research who approached CHET using email with data queries was done via email correspondence in an unstructured manner.

Section 3. Findings: Technical platforms and standards

CHET

CHET is committed to using open source software where possible. All project management, development and maintenance of the website is outsourced.

The CHET website is built using Drupal 7 – an open source web content management solution. The open data platform resides within the CHET website architecture. The open data platform has undergone three major transitions since its launch on 13 November 2009. In the first phase, the CHET website simply linked to an online graphing tool that allowed users to generate custom queries using the CHET indicators populated with data from HEMIS. The graphing tool allowed users to generate on-the-fly graphs and to download the data related to their specific query as an MSEXcel file. Phase 2 saw the creation of a dedicated sub-section for CHET's open data. During this phase, additional datasets were made available for download as csv files and as MSEXcel files. Included was the full indicator dataset for download as either an archived file consisting of individual csv files for each indicator or as a consolidated MSEXcel file containing all the indicator data. The primary change during Phase 3 was to create two separate sections for data related to South African higher education and data related to African higher education.

The technology stack used for the development of the graphing tool (<http://chet.org.za/indicators/>) is as follows:

- PHP scripting, Smarty template engine, MySQL database, jQuery.
- To generate the graphs Google Visualization is being used.²⁵
- To enable users to download data results as MSEXcel files, ExcelExport is being used.

The CHET data platform displays the logo of and links to the Open Data Commons. Open data license is clearly indicate in the footer of the open data platform: "CHET data and databases are made available under the Open Data Commons Attribution License: <http://opendatacommons.org/licenses/by/1.0>". Similar licensing information is displayed on each user-generated graph and is included in the downloadable csv and MSEXcel files.

HEMIS

As a result of the 1997 White Paper on Higher Education, a system for collecting unit record statistical data from all universities was developed and implemented in 2000. The data collected comprise student data (three times per annum), staff data (twice per annum) and building space data (once per annum); it does not contain any financial data.

HEMIS is a SQL database hosted and managed by DHET but currently maintained by a private IT company, Praxis. The system requires universities to:

1. Prepare data from their databases as ASCII files.
2. Import the data into an Access database within a PC application "Valpac" which is provided to them by DHET. Valpac generates the Access database. The application functions on stand-alone single-user PCs within each institution.
3. Use Valpac to validate the data and to generate comprehensive reports, which are used by the institution in quality assurance. Valpac has functionality for validating the data for generating reports.

²⁵ See <https://developers.google.com/chart/interactive/docs/reference>. An earlier version of the graphing tool used JpGraph PHP5.

4. When the quality of the data has been assured, universities use Valpac to create a zipped and encrypted database, which is then sent to DHET via email (or, in the case of the University of South Africa [Unisa], via ftp due the file size of the database).

When the Department receives the zipped and encrypted Access databases from each institution, they are inspected using Valpac and the validation processes are replicated by Departmental staff and the reports produced, using Valpac functionality. The data from the databases are then loaded into a server-based application called “HEMIS”. This application creates a national database from the individual returns from each institution. It is used to generate comprehensive reports for selected institutions, groups of institutions, and for all institutions combined. Tables in the national database are also useable for ad hoc queries using Access. The HEMIS application functions across a LAN.

The Valpac and HEMIS systems were developed in Visual Basic .Net with the latest service pack and using the MS Access 2007-compatible Microsoft Jet engine and MS SQL 2005 respectively as their databases.

Section 4. Findings: the supply of (open) data

There has undoubtedly been phenomenal growth in the supply of government open data. Davies et al. (2013) estimate that governments have posted in excess of 1 million datasets online. As at June 2013, Datahub counted 6588 datasets (although not all of them are open datasets). In terms of open data, DataCatalogs.org counted 337 open data catalogues and portals worldwide each consisting of several datasets and the Linked Open Data Cloud counted 295 linked open datasets. This section of the report seeks to capture how open data is supplied in the very specific context of South African public university governance, and what the flow of data reveals about the dynamics of open data supply.

The flow of data, both non-public and open, in the South African public university system is shown in Figure 1. The representation of the flow of data within institutions has been simplified although the interviews with university planners show that all institutions in the sample have a management information system (MIS) in place to collect institution-wide data from the various university academic and administrative units. Data input into the MIS varies along a continuum from being decentralised to centralised. The process of data extraction from the MIS, data validation and data uploading to HEMIS (via a private IT company appointed on contractual basis by DHET) appears generic to all public universities in South Africa. In all cases, universities had dedicated administrative personnel responsible for collecting, preparing, validating and submitting data to DHET.

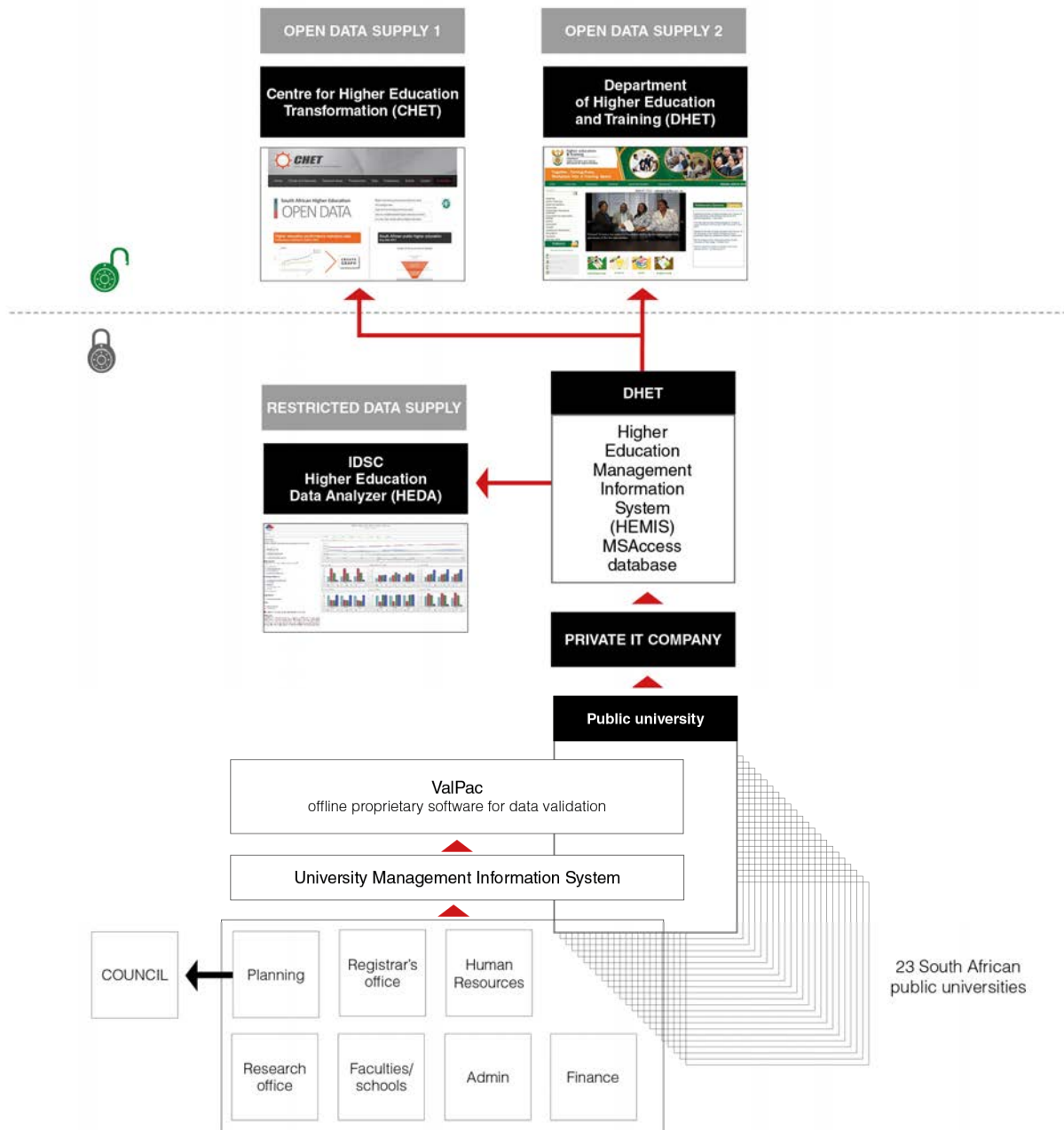
HEMIS is a SQL database hosted and managed by DHET but currently maintained by a private IT company, Praxis. The database contains student and staff unit records, as well as building space data; it does not contain any financial data. Direct public access to the data is restricted as the unit records contain personal data which would contravene privacy rights should the data be made public. For a third party to be granted full access to the database, the vice-chancellor from each university for which data has been requested is required to grant authorization. Within DHET, only the Director of HEMIS, her two deputies and the assistant director have full access to the database.

From the HEMIS database, two *open* datasets are made public on the Internet. The first of these is the DHET open dataset, which is available on the DHET website.²⁶ DHET publishes eight *anonymised* data tables (on enrolments, graduates and staff) in Microsoft Excel format on its website. The second open dataset is supplied by a non-governmental organisation, the Centre for Higher Education Transformation (CHET): <http://www.chet.org.za/data>. CHET publishes 18 data tables (each related to a specific performance indicator) in two formats: (i) as downloadable csv files and (ii) through an interface which allows users to generate custom graphs and data tables per indicator with the possibility of comparisons across a maximum of four universities.

A notable dimension of the CHET open data is how the source data is obtained. Initially, the source data could only be obtained on request to DHET. However, since the publication of the anonymised HEMIS data on the DHET website, the data can be accessed directly. Despite this, CHET employs a consultant who has worked with DHET for a number of years to source the data from DHET. In other words, while CHET could extract the data it requires from an open data source, in practice the process of data transfer from a government department to an external entity is facilitated by the professional relationship between the a consultant and key personnel in a government department.

²⁶ <http://www.dhet.gov.za/Structure/Universities/ManagementandInformationSystems/tabid/419/Default.asp>

Figure 2: The flow of data in South African higher education



A third data provider extracts data from HEMIS. The Higher Education Data Analyzer (HEDA) is a product of the private IT-company, IDSC (see <http://www.heda.co.za>). HEDA is excluded from the discussion in this section of the report on the supply of open data because, based on the Open Definition, it is not considered to be an open dataset. HEDA's terms of use read as follows: "The Department of Higher Education has requested that all PDS users agree to the following: The data provided by the PDS website

will only be for internal use and not for external publication.” The role of HEDA is, however, discussed later in this report as its presence cannot be ignored in the higher education governance data ecosystem. What is clear from a supply perspective, is that HEDA sources data from the same source as CHET, but provides restricted rather than open access or re-use.

Both CHET and IDSC play an intermediating role in the flow of data – they are positioned between the primary data source (HEMIS) and the market. *How* CHET and IDSC access and provide data from the same data source raises further questions to be explored around parallel public and private data flows, the role of social capital and trust in accessing source data and the unequal provision of (re)use rights from a common government data source. Intarakummerd and Chaoroenporn (2013) in their research on the role of intermediaries in innovation in developing countries, highlight the role of intermediaries in compensating for a lack of social capital in innovation systems. They also point to the importance of government initiating and coordinating the activities of both public and private intermediaries.

The fact that two *open* datasets exist from the same source also raises the basic question of why this apparent duplication of open data provision exists in the first place. The fact that DHET provides open HEMIS data could be attributed to a government-wide pledge to open data provision in order to validate its commitment to transparency and accountability (South Africa was a founding member of the Open Government Partnership launched in 2011). However, based on the interview conducted with DHET, the more likely reason for the provision of open data is to deflect those who approach DHET for HEMIS data to the online open data tables. DHET has limited capacity to deal with requests for data – only four DHET staff have access to the full dataset – and sharing the dataset online is therefore an attempt to reduce the burden placed on the Department by external requests for data. The fact that the dataset is difficult to locate on the DHET site seems to support this finding from the interview. If the motivation for opening up the data was transparency or (re)use, one would expect it to be much easier to locate the data tables on the DHET website. However, if the motivation is simply the ability to redirect data queries, then all that is needed is a hyperlink to the data (however obscure that link may be) which can easily be emailed to those directing requests at DHET staff.

In the case of CHET, the supply of open data is premised on a clearly identified governance need: “The requirement that each higher education institution must confirm its acceptance of planning targets makes it essential that councils understand (a) what is implied by the targets, and (b) how their institution is performing relative to these targets. CHET’s experience has been that this has not been an easy task for councils. Currently very few institutions produce datasets which would enable council members to engage meaningfully in discussions about the performance of the institution which they are entrusted to govern. CHET decided [...] to produce data profiles which should enable university councils to make assessments of the performance of their institution relative to the targets” (Bunting et al., 2010).

The reason for the dual provision of open data on South African public universities seems to be that each supplier is driven by different priorities. In the case of DHET, the supply is driven by internal factors – a lack of capacity – and with no particular reference to users and what their data needs may be. A government-wide commitment to improved system-level governance through transparency and accountability has no bearing on the provision of open data. In the case of CHET, supply is driven by a perceived need for improved governance at institutional level in the light of government-set targets through evidence-based decision-making in relation to such targets. Supply is therefore very much focused on the user in CHET’s case.

It is clear that CHET did not, however, intend for its performance indicator data to be used by higher education studies researchers when conceptualising and developing the indicators. However, this does not necessarily imply that CHET’s open data platform was designed with a singular audience in mind. In fact, a statement on the CHET open data platform reads as follows: “CHET has [...] decided to make

performance indicator data available on-line to enable university councils to better assess their performance relative to the Minister's targets, to their own institutional targets and to the performance of their peers. This data tool will also be of use to higher education researchers, analysts, policy-makers and other decision-makers seeking a more detailed, empirically-based picture of South African higher education." In the section of this report on the use of the CHET open data platform, consideration will be given to whether the platform meets the needs of both target audiences. What is of importance here, is that both of the target audiences are considered by CHET to have roles to play in improving the governance of South African public higher education (be it at the system or institutional level) by means of the data supplied by CHET on its platform.

Underlying the supply of CHET open data in public university governance is a theory of change about how open data may affect governance. CHET's theory of change is that evidence-based decision-making will improve the governance of public universities. Returning to the basic definition of governance being concerned with processes of decision-making and implementation, CHET's ambition to effect change through better decision-making seems reasonable. In the case of DHET, there is no apparent theory of change. The data is made available for internal reasons without any consideration of the user and how they might (re)use the data. These differences affect how each organisation supplies the data. CHET provides 'shaped' data in the form of indicators because it believes that it will have the greatest chance of effecting change in this format. DHET, on the other hand, effectively 'dumps' data on its website with little by way of contextual information to guide the user on how to use or interpret the data. That the data is as complete as possible is important to DHET as it increases the likelihood of the dataset covering the full range of possible types of data requested by those approaching DHET for data.

Further differentiation on the supply side between the open datasets is possible based on the extent to which datasets can be regarded as open. Multiple definitions exist around the core concept of the Open Definition. Different interest groups place varying emphasis on specific requirements depending on the context in which they anticipate the use of open data to take place. The Exploring the Emerging Impacts of Open Data in Developing Countries (ODDC) project has developed a set of 10 criteria against which the openness of a dataset may be assessed. A 2007 working group comprised of open government advocates developed the 8 Principles of Open Government Data against which it suggests open government data should be assessed.

CHET's open data scores better on the ODDC criteria [9/10] than the DHET open data [7/10] because it is openly licensed and easier to locate (see Table 1). That the CHET data is both licensed and easier to locate than the DHET open data confirms a more user-orientated approach on the part of CHET. Again the scores are different when the datasets are assessed according to the Open Government criteria (see Table 2). The reason for different scores in the case of the Open Government principles is that the DHET open dataset complies with the principle of the data being timely, complete and primary (albeit that the DHET data is not strictly primary but an anonymised version of the primary data). In the case of the CHET open data, the relevant data is extracted from HEMIS and supplied as indicators. As intermediary, timeliness is a condition over which CHET can never exercise full control unless DHET were to provide real-time access to the HEMIS database. CHET will always be dependent on the release of data by DHET, and will lag behind DHET in the release of open data. Compounding the timely release of data by CHET are the resources (both financial and human) required to convert the raw HEMIS data into the indicator dataset.

Table 1. Evaluation of openness of two datasets using the Exploring the Emerging Impacts of Open Data in Developing Countries 10-point evaluation

ODDC OPEN DATA CRITERIA	DHET	CHET
Does the data exist?	✓	✓
Is it available online in digital form?	✓	✓
Is the data machine readable?	✓	✓
Is the data available in bulk?	✓	✓
Is the dataset available free of charge?	✓	✓
Is the data openly licensed?	✗	✓
Is the dataset up-to-date?	✓	✓
Is the publication of the dataset sustainable?	✓	✓
Was it easy to find information on the dataset?	✗	✓
Are linked data URIs provided?	✗	✗
TOTAL SCORE	7	9

Table 2. Evaluation of openness of two datasets using the 8 principles of Open Government Data

OPEN GOVERNMENT DATA CRITERIA	DHET	CHET
Data must be complete	✓	✗
Data must be primary	✓	✗
Data must be timely	✓	✗
Data must be accessible	✓	✓
Data must be machine processable	✓	✓
Access must be non-discriminatory	✓	✓
Data formats must be non-proprietary	✗	✓
Data must be license-free	✗	✓
TOTAL SCORE	6	5

From the scores based on the criteria set out by the two open data definitions, it is clear that both datasets can be considered to be open and that the two open datasets are different in fundamental ways. They are different (i) because of the implicit differences inherent in being a primary data provider and being an intermediary provider, and (ii) because each provider has different motivations for opening up the HEMIS data. However, Davies (2013) cautions: “scoring highly [...] does not mean that the dataset in question is necessarily useful to anyone.” Having provided some insights into the dual supply of open data in the governance of South African public universities, attention now turns to this critical issue of the use and usefulness of the data provided to university planners and higher education studies researchers.

Section 5. Findings: data use and impacts

Use of CHET open data by university planners

The tables below set out the findings from the interviews conducted with the unit responsible for data collection, data management and/or data requests at the universities in the sample. Typically this organisational sub-unit was the Institutional Planning department located within the larger university administration.

Before the findings on the use of the CHET open data are presented, consideration is given to the origin and nature of the data requests directed at university planners. The intent is to provide a glimpse of the user context.

Table 3 indicates that requests for data directed at university planning departments are both external and internal. Universities are approached directly by a range of external groupings interested in university data. These external requests indicate that DHET is not the sole provider of data even though it is the central source of higher education data. Internal requests are varied across the institution. Most of the requests come from deans of faculty, heads of department or other academic staff. The table indicates that NMMU is the exception to this trend. The reason for this is that, in the case of NMMU, requests for data are automated through the provision of a centralised data system by the university that allows academic staff to run data queries themselves. UP has a similar system although requests are often still made directly to the planning staff (either telephonically or via email). At other universities in the sample, data requests are by and large manual.

Table 3: To university planners: From whom do you receive requests for data?

University	EXTERNAL				INTERNAL					Students
	Bus.	Gov.	Media	Research	Council	Exec.	Faculty	Registrar	Other units / committees	
DUT						x	x			
NMMU					x	x				
UCT				x		x	x		x	
UFH		x	x	x	x		x			x
UJ			x	x		x	x	x	x	
UP	x	x	x	x		x	x			
WSU							x			
TOTAL	1	2	3	4	2	5	6	1	2	1

Few requests emanate directly from the university councils. This is most likely due to the fact that requests for data are channelled via the executive members of council who make up the “Executive” category in the table. A large number of requests originate from the university executive. Exceptions are the UFH and WSU. In the case of UFH, requests originate directly from the university council, although requests are usually limited (i.e. only for data on student enrolments). In both interviews, it was confirmed that data requests from university leadership were infrequent.²⁷ In fact, it appeared that at WSU and UFH, the university planners were pushing data up to the university leadership rather than assuming a more responsive role as is the case at the other universities in the sample. At WSU, the data presented to

²⁷ According to the interviewee at WSU, a lack of data use at executive levels results in very little downstream pressure in assuring accuracy in data capturing. This could indicate a vicious cycle of poor data input and poor data uptake.

council was often met with suspicion and with an insistence that the data be verified by a third party before being considered by council as reflecting an accurate state of affairs. Coincidentally, WSU is the only university in the sample where there are known governance issues – the university is one of five currently under administration.

Table 4 shows that, based on the interviews conducted, data requests are generally still relatively basic. Few respondents indicated that they receive requests for data other than enrolment and success rate data. There appeared to be a general sense that limited data literacy at council level is at the heart of the relatively unsophisticated nature of the data requests.

Table 4: To university planners: What kind of data is typically requested?

University	Student enrolments	Success rates	Research outputs	Staff	Financial	Funding
DUT	x	x	x			x
NMMU	x	x				x
UCT						
UFH	x					
UJ	x	x				
UP		x				
WSU	x	x				
TOTAL	5	5	1	0	0	2

Table 5 shows that all but one of the interviewees was aware of the CHET open data platform. However, Table 6 shows that use of the platform was infrequent. The infrequent use could be attributed to infrequent updates as indicated in the “Improvements” column in Table 7. HEMIS data is only released in full annually so the most frequent update possible is once per annum. However, at the time of the interviews, the CHET data open platform had not been updated for two years. Other areas for improvement expressed by the interviewees and which could be interpreted as being possible reasons for the infrequent use of the CHET open data are a lack of a broader set of indicators (such as throughput data) and/or data that conforms with the indicators used by the global university rankings agencies.

The most useful feature of the open data platform was the ability to access comparative data. While access to their own data is a given, the university planners do not have access to the data from other universities. Their only recourse other than the CHET open data platform is to approach DHET for comparative data, or to access the data on the DHET website.

Table 5: To university planners: Are you aware of the CHET open data platform?

University	YES	NO
Durban University of Technology (DUT)	x	
Nelson Mandela Metropolitan University (NMMU)	x	
University of Cape Town (UCT)	x	
University of Fort Hare (UFH)	x	
University of Johannesburg (UJ)	x	
University of Pretoria (UP)	x	
Walter Sisulu University (WSU)		x
TOTAL	6	1

Table 6: To university planners: Have you made use of the CHET open data to support the governance of the university?

University	Yes, frequently	Yes, infrequently	No, but likely to do so	No
Durban University of Technology			X	
Nelson Mandela Metropolitan		X		
University of Cape Town (UCT)		X		
University of Fort Hare (UFH)		X		
University of Johannesburg (UJ)	X			
University of Pretoria (UP)		X		
Walter Sisulu University (WSU)				X
TOTAL	1	4	1	1

Table 7: To university planners: How is the open data platform useful to you? What changes would you like to see?

University	Useful	Improvements
DUT	n/a	n/a
NMMU	n/a	Provide second-level categories for SET data Comparisons against ministerial priority areas Fewer financial indicators
UCT	Comparative data	Greater granularity; no ratios or percentages Increase the number of possible comparisons to eight
UFH	Comparative data Values and ratios are used when universities are compared	More timeous updates of the HEMIS data
UJ	Student:staff ratios Provision of both FTE and head count data	Indicators similar to those used in rankings More timeous updates of the HEMIS data
UP	Provides triangulation with other data sources Comparative data	Indicators similar to those used in rankings
WSU		Throughput data

When asked other data sources they rely on, university planners indicated that they source data from the CHET open data platform, from the Higher Education Data Analyser (HEDA) platform and directly from DHET (see Table 8). Requests to DHET were usually made in person to one of four DHET staff who have access to HEMIS; only one interviewee made use of the DHET data tables published on the DHET site. Three of the seven universities cited HEDA²⁸ as their most frequent source of data. At the time of writing, 12 of the 23 public universities in South Africa were making use of HEDA.

²⁸ <http://www.idsc.co.za/>. See <https://heda.cput.ac.za/indicatordashboard/default.aspx> for an example of a data dashboard that a university has elected to share publicly.

Table 8: Which other external data sources do you use? Which do you rely on the most?

University	CHET	HEDA*	DHET	OTHER	Used most often
DUT				x	OTHER
NMMU			x		DHET
UCT	x	x	x		HEDA
UFH	x		x		DHET
UJ	x	x			HEDA
UP	x	x	x		=
WSU		x	x		HEDA
TOTAL	4	4	5	1	

* Higher Education Data Analyser

= No predominant external data source

Use of CHET open data by higher education studies researchers

The data collected in this research on those using open data for the purposes of higher education studies research can be split into two groups: (i) Interview transcripts and (ii) transcripts of emails submitted via CHET open data website and requesting data. The data collected from the two groups differed due to the nature of the data collection process followed. The data extracted from these two groups are presented in the tables below. (Data collected from the email transcripts is indicated by means of an asterisk in the tables below.)

From the interviews conducted, all the researchers are aware of the CHET open data platform and make use of the data (see Table 9). Table 10 shows that other commonly used data sources include Unesco and the World Bank. Half of the higher education studies researchers interviewed make use of the open data tables on the DHET website.

**Table 9: To higher education studies researchers:
Are you aware of the CHET open data platform?**

Researcher*	YES	NO
QQ	x	
HP	x	
RG	x	
KC	x	
UK	x	
HI	x	
TOTAL	6	0

* Anonymised initials used to protect the identities of those interviewed

Table 10: To higher education studies researchers: Which online open data sources do you use?

Researcher	CHET	World Bank	Open Data for Africa	DHET	HEDA	OTHER
QQ	Y	Y	N	N	N	NIDS (UCT) UNDP Unesco
HP	Y	Y	N	N	N	OECD
RG	Y	N	N	N	Y	--
KC	Y	Y	N	Y	N	Unesco
UK	Y	N	N	Y	N	AfroBarometer, Freedom House
HI	Y	Y	N	Y	N	--
TOTAL	6	4	0	3	1	

Table 11 indicates that the most useful features of the CHET open data platform is the provision of university-level data (sources such as the World Bank only provide system-level data) and its ease-of-use. Commonly expressed areas for improvement include the provision of more data and changes to the functionality of the platform to allow for more complex analyses of the data.

Table 11: To higher education studies researchers: If you use the CHET open data platform, how is it useful to you? How could the platform be improved?

Researcher	Useful	Improvements
QQ	Provides institutional-level data while most other sites provide system-level data	Richer data. Data included from other CHET research projects.
HP	Provides data that is credible in the eyes of the university executives Provides institutional-level data	--
RG	CHET platform is useful in providing comparative analysis.	Can be improved by comparing more than four institutions. Maybe there could be an option to use a cross tabulation which could bring in more than one variable.
KC	Very useful indeed and easy to access and use	Nothing that crosses my mind right now -- more data perhaps?
UK	--	If I remember right (it's been a while that I used CHET indicator data), I would have liked to be able to download the data sets as a whole; I think I could only download analysis.
HI	I find the HE institutional data very useful for contextualising purposes	There should be a facility to be able to design your own tables for analysis = the current range is too restricted.

Table 12: Types of data requested by those who submitted data requests via the CHET open data platform

Researcher*	Data requested	CHET able to supply: Y/N
UN	Comparison of enrolments by major field of study	N
QS	I'm primarily interested to see if and how white English and white Afrikaans students migrate between different South African universities, but migrations of coloured, Indian and black students would be interesting too. It would also be important to be able to break down institutions into their different campuses. I'd want to see race and language stats per university per campus per level of study.	N
FN	We would like to do a more in depth analysis on higher education than what the current data will allow us to do. The data will be used for some research we are doing on upward mobility. 1. Enrolment numbers for new students for both first years and postgraduate students by race 2. Enrolment by specific field/degree (not just major field) by race 3. Enrolment for new first year students by specific field/degree (not just major field) by race 4. Actual graduates (for undergrad and postgrad) by race 5. Is there any indication of average grades for students? If so, could we see this by university and by race? 6. Cohort analysis: is there any way to track a specific group of students over time?	N
RL	Graduate throughput / success rates	Y
KD	I would like the UCT-specific and national sources of income	Y
LH	To perform analysis that will be used in a publication	Y

* Anonymised initials used to protect the identities of those interviewed

The data requests submitted via the CHET website confirm that, in general, higher education studies researchers want access to more granular data (see Table 12).

Those who made data request via the CHET data platform, are generally unaware of the DHET as a source of data (see Table 13). As confirmed by follow-up emails, they found it difficult to locate the correct person(s) in the department when directed to DHET. Those who are aware of the DHET as a source of data tend to access the data by requesting the data directly from DHET, while those who use the open data tables are in the minority, and find using the data tables frustrating.

Table 13: Are you aware of the DHET open data tables?

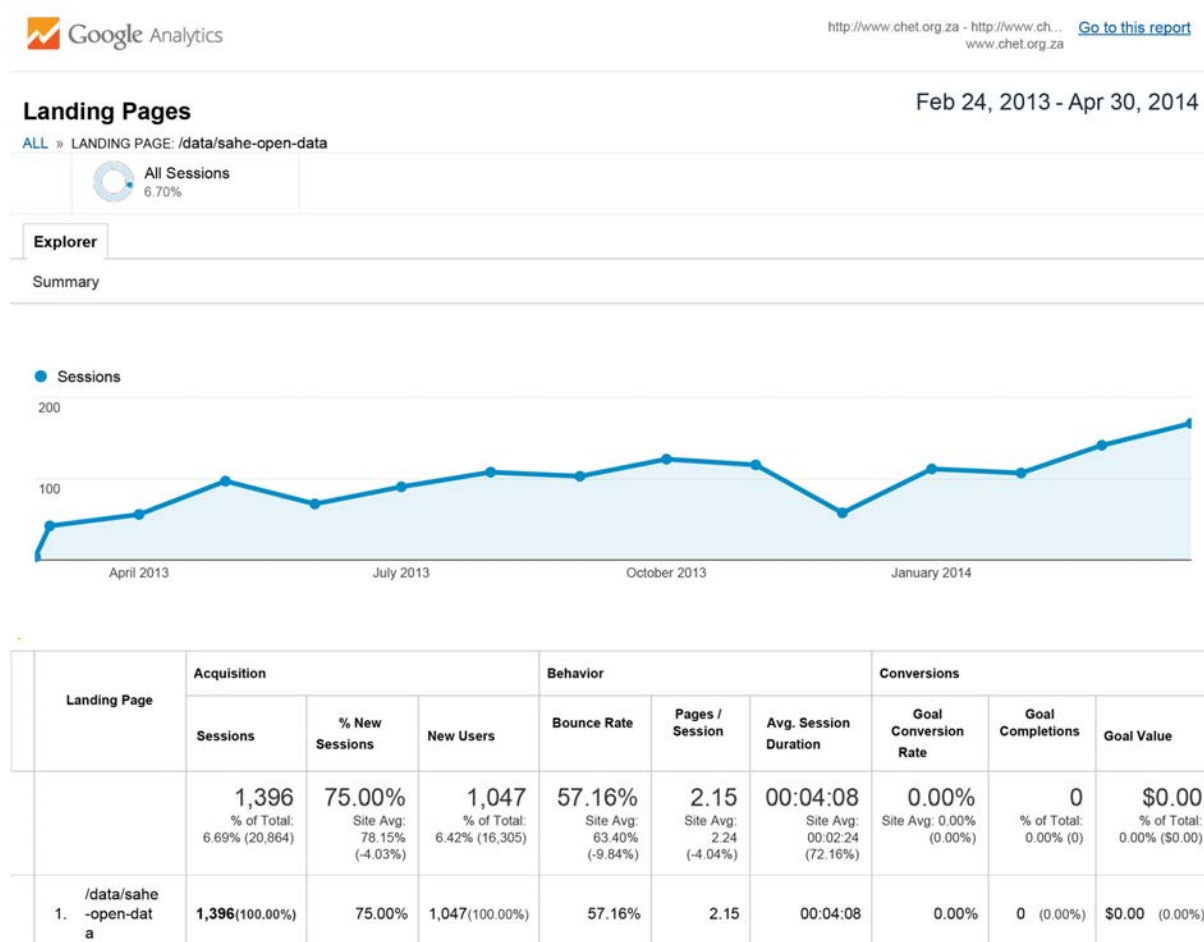
Researcher	YES	NO	If YES, how do access the DHET data? Online data tables / Request by phone or email / Both
QQ	x		Request by phone or email
HP	x		Request by phone or email
RG	x		Access HEMIS data via HEDA, using someone else's login details
KC	x		Online data tables; a frustrating process
UK	x		Online data tables
HI	x		Request by phone or email
UN		x	
QS		x	
FN		x	
RL	--	--	
KD		x	
LH	--	--	
TOTAL	4	4	

Usage based on web statistics

The ODDC project had access to CHET's website statistics generated by Google Analytics. This data provides an additional indication as to whether the CHET open data is being used, although without knowing exactly who is using the data and how they are using it, the link between usage and university governance remains extremely tenuous.

An analysis of the CHET web statistics for the period 24 February 2013 to 30 April 2014²⁹ shows that the South African higher education data page was the second most popular landing page on the CHET website (after the home page). The recorded bounce rate of 57% is relative high, indicating that over half the visitors to the open data section make no use of it. This can be explained by the fact that 71% of traffic to the open data section of the website originates from search engines (Google 68%, Bing 2% and Yahoo 1%), increasing the likelihood of mismatch between user expectation and content served. Despite the high bounce rate, the average session duration is just over 4 minutes. This would seem to indicate that those users who remain on the open data platform, engage substantively with the content (data) provided.

Figure 3: Selected statistics for the CHET open data platform



²⁹ The date range covers Phase 3 of the CHET open data platform (see section on Technology in this report). This period provided the most accurate data related exclusively to the use of South African higher education open data.

User onus and data marketisation

A footnote on data use: The CHET open data platform includes functionality that allows users to submit general comments on the data platform as well as specific comments related to any data query generated on the platform. A typical user comment is received by email and is set out as follows:

Name	EVN
Email Address	xxx@nwu.ac.za
Graph Url	http://www.chet.org.za/indicators/indicator.php?uid=52&aid=&iid=36&rid=123&gw=800
Date and Time	2012-01-24 15:59:56
Data Errors	I am from NWU and took over from ADW; According to her and NWU management the \"Headcount enrolments\" of NWU for 2000 are incorrect. It should be 24.0 and not 23.6 as indicated in the graph. I'll appreciate it if you can adjust the figures. EVN: Manager MIS

CHET has received no more than three such comments since the platform was launched in 2009. This could indicate that users have very few or no complaints about the functionality of the platform, the breadth of the data contained on the platform, or the accuracy of the data on the site. However, the interviews conducted indicate that there is room for improvement, particularly in relation to functionality and the breadth of the data. This points to some onus on the part of users – if users want an improved open data platform that better serves their needs, they need to communicate their needs to CHET. Of course, users are always free to simply choose an alternative service if their needs are not being met. In the case of open data on higher education, with the exception of DHET, there are currently no other open data providers. This situation could change, and CHET therefore needs to remain mindful of the needs of its target audience. In fact, the emergence of the subscription-based Higher Education Data Analyzer (HEDA) could be seen as a move into this space, and one that makes CHET's role as provider of open data all the more critical as higher education data becomes marketised.

Conclusion

CHET's open data is being used by university planners and by higher education studies researchers, albeit infrequently. University planners found CHET's performance indicator data useful and some planners expressed interest in additional indicators being made available as open data. Researchers expressed the need for richer, more granular data. Both planners and researchers expressed the value of the comparative, institution-level data made available by CHET.

This finding may prove useful should CHET be planning any modifications to its open data platform. CHET may need to give consideration to focusing on the needs of university planners over those of higher education researchers as such an approach is more closely aligned with its original strategy for providing open data, that is, to strengthen the governance capacity of public universities. Also, CHET may find that it does not have the capacity or resources to share large volumes of granular open data; a task that is best left to DHET. Either way, CHET may need to carve a more differentiated role for itself as a supplier of open data based the needs of different user groups and given the presence of other more financially secure intermediaries such as IDSC who have already secured a foothold in the market.

University planners at the universities interviewed indicated that a large number of data requests originate from the university executive who require data for strategic decision-making and/or for presentation at council meetings. Exceptions are the UFH and WSU, where the analysis and interpretation of data is done by the university planners and “pushed upwards” in the decision-making hierarchy. This could point to an uneven distribution of data literacy and possibly uneven levels of data trust at council level in South African universities. The uptake and impact of open data on university governance could be hindered at

universities at which low levels of data literacy and/or mistrust prevail. Given that WSU is also under administration, one may be tempted to draw links between poor governance and low levels of empirically-based decision-making at executive and council levels. This study confirms such a link in the case of WSU but further research is required in order to make any generalisable conclusions.

In addition to mistrust at university council level, there are concerns at both government and university planning levels about how their data will be used and (mis)interpreted. This may constrain future data supply. Education both at the level of supply (DHET) and at the level of use by the media in particular on how to improve the interpretability of data could go some way in countering current levels of mistrust.

HEMIS is a closed and isolated data source; and our findings show that DHET is often not the first port of call for data users. Granting access to HEMIS by third-parties (such as CHET) under controlled conditions to protect personal data could further stimulate the provision of open data and relieve pressure on the capacity-constrained government department. This could bolster the impact of open data on the governance of South African public universities. Discussions between DHET, CHET and other stakeholders on how to share HEMIS data and how to improve the interpretability of the current DHET open data tables, could stimulate a new phase in the evolution of the higher education governance open data ecosystem.

It is to this ecosystem to which we now turn our attention in order to understand CHET's role in the broader dynamics of the supply and use of university governance open data.

Section 6. Discussion: The higher education governance data ecosystem – data viscosity and role of intermediaries

Yu and Robinson (2012) describe open data as being either “adaptable” or “inert”. Manyika and colleagues, in their recent work on quantifying the economic value of opening up data, alight on the notion of open data being “liquid” (Manyika et al. 2013). That is, open data unlocks value as it flows from governments, between firms, researchers and entrepreneurs, and to citizens, and is adapted in the process. To extend the analogy borrowed from the natural sciences, the flow of data could result in a virtuous cycle, becoming a stable but dynamic part of an ecosystem. But equally possible, data could, despite being open, become inert and flow too slowly or not at all; it could be too viscous to contribute to the evolution of the ecosystem.

As business intelligence becomes entrenched in university planning practice (as a basis for strategic decision-making in a marketised system), data is likely to become increasingly proprietary rather than open at institutional level. This, it could be argued, places greater importance on provision of open data by government; and it accentuates the importance of sustainable open data solutions by intermediaries.

At the same time, the government’s Department of Higher Education and Training (DHET) whose responsibility it is to manage the central higher education management information system (HEMIS) database as well as some of the university planners interviewed, expressed concerns around how and by whom HEMIS data is accessed. They expressed concerns about the misrepresentation of South African higher education due to poor data literacy on the part of the public and, in particular, the media.³⁰

Both of these conditions – data for the use of competitive advantage and user mistrust – have the potential to constrict the supply of open data.

The case of the South African public university open data is highly relevant because of the presence of intermediaries in the supply of open data and the concomitant existence of multiple open data resources from a single, closed government data resource, in a capacity-constrained environment. This suggests that the selection of this case could reveal potentially valuable insights into the dynamics of a complex open data ecosystem.

This section of the report is therefore concerned with the flow of data to and from the government HEMIS database, how the provision of open data by the government is contributing to the evolution of the ecosystem and whether the presence of those intermediating between the provision and use of data are contributing to the ecosystem. In addition, the presence of intermediaries in the ecosystem is explored in an attempt to assess their contribution to the fluidity of data and, ultimately, to the evolution of the ecosystem. In ‘constructing’ this particular ecosystem, attention will be directed towards the actors in the ecosystem, their roles as providers or consumers of data, and how they are inter-related. Several questions will be answered: What/who are the key drivers in the system? What is the role of government within the system? Who are the intermediaries in the system, and what role do they play? How do the modalities of open data supply impact the ecosystem? Who or what are the keystone species (drivers and enablers) present in the ecosystem?

The conceptual framework for this part of the report borrows heavily from Helbig et al.’s (2012) ‘information polity’ heuristic. In their white paper *The Dynamics of Opening Government Data* they propose the heuristic device of an ‘information polity’ to model and analyse the context and dynamics of open data initiatives. Included in their information polity are the actors and their roles, data sources and

³⁰ This finding is consistent with that of a study by the Royal Statistical Society in which only 4% of UK adults indicated that they trusted the media to use their data appropriately. http://www.statslife.org.uk/files/perceptions_of_data_privacy_charts_slides.pdf

flows, and the governance relationships in open data initiatives. We use their information polity as a starting point in modelling the open data ecosystem of our case. In particular, the definitions provided and distinctions made between primary and secondary open data sources, resources, providers and users are relied on to chart the relationships between these components of the information polity. However, we extend the information polity heuristic into a data ecosystem partly because we believe that the concept of an ecosystem enables the more accurate reflection of the resources, sources, providers and users in a context broader than when government alone acts as the primary collector and provider of data, and partly because we believe that the concept of the ecosystem will resonate to a greater degree with both practitioners and scholars, both of whom we hope will find value in the mapping and analysis of contexts and dynamics in the provision of open data.

The concept of the ecosystem has already gained a degree of traction in the analysis of how ICTs are driving change, be this in discussions on open government or open data. Harrison et al. (2012), in a review of the ecosystem metaphor in the open government literature, identify several key features of ecosystems. Ecosystems are seen as consisting of mutually interacting organisms; complex in their arrangement; characterised by the interdependency of and between organisms and resources; dynamic rather than static – seeking equilibrium through motion rather than stasis; populated by keystone species that play a critical role in facilitating exchange in the ecosystem thereby ensuring dynamism and constant movement; movement tends to be cyclical and reinforcing making the system resilient (adaptable and restorative); but it is also vulnerable to exogenous forces which may disrupt or destroy the ecosystem.

Martin Fransman, in his book *The New ICT Ecosystem*, draws on the work of evolutionary economist Joseph Schumpeter, to describe the components of socio-economic ecosystems and to recast these components in the context of ICT which he argues constitutes one of many sectorial ecosystems within the larger socio-economic ecosystem. He identifies the dynamically interacting organisms in the ICT ecosystem (firms, non-firms, intermediaries and consumers) bound by exchange as well as by the institutions (the repositories of rules, values and norms) in which they are embedded. Key to his exposition of the ICT ecosystem is that the ICT ecosystem is driven by innovation (i.e. the injection of new knowledge into the ecosystem). Firms compete and co-operate symbiotically, and the interaction between firms and consumers (that is, between knowledge creators and knowledge consumers) generates new knowledge which leads to innovation in the ecosystem.³¹ It is the pursuit of innovation that keeps the ICT ecosystem in motion.

For the purposes of our analysis of a particular data ecosystem: If knowledge creation as a simplified process moves from observation to recording those observations to analysis to testing to validation, and data is the codified retrievable recording of observations in this process of knowledge creation, then it seems reasonable to assume that the open data ecosystem is a key component in the broader ICT ecosystem, particularly if it is premised on innovation as a key driver. What is less clear is whether innovation per se is a driver in an open data ecosystem or, if it is a driver in the open data ecosystem, what conditions need to be in place to ensure the sustainability of such an innovation-driven ecosystem.

Finally, we extend the information polity framework by adding the concept of “keystone species” which are considered crucial to ecological functioning because their presence performs some vital function (Nardi and O’Day 1999: 53). These enabling actors in the ecosystem could take the form of mediators, actors who bridge institutional boundaries and translate across disciplines, or they could “create value for their ecosystems in numerous ways, but the first requirement usually involves the creation of a platform,

³¹ Muller and Cloete (1986) contend that knowledge validation is the exclusive preserve of academia, while firms apply validated knowledge to innovate (that is, to create or improve new systems, process, products, organisations). It is beyond the scope of this paper to engage in detail with the literature on the sociology of knowledge. We do, however, include both firms and non-firms in our open data ecosystem.

an asset in the form of services, tools, or technologies that offers solutions to others in the ecosystem” (Iansiti and Levin 2004: 7).

Actors

University planners were identified as key actors in the flow of data and as potential users of open data in fulfilling their task of advising the university executive and council in the process of strategic planning. In order to extend the demand-side contours for open data in the public university governance ecosystem, the project identified higher education studies researchers as secondary users of such data. From a governance perspective, higher education studies researchers play a key role in the research–policy nexus. Higher education researchers are located in universities, research NGOs, government agencies and consultancies. Finally, the Department of Higher Education and Training (DHET) as the sole aggregator and supplier of primary system-wide, institutional-level data was included as an obvious key actor in the public university governance ecosystem.

The study did not include citizens or the media in its analysis. It could be argued that citizens (including students and student organisations) and the media are able to play a role in the governance of South African higher education by holding universities and government, as the custodian of higher education, accountable. However, this study’s focus on open data as a mechanism for improved governance through informed, empirically-driven decision-making, rather than accountability, limited the analysis on the demand-side to university planners and higher education studies researchers. This is not to suggest that accountability and informed decision-making are not linked or mutually reinforcing.

Open data sources

The identification of data sources was done by means of desk research combined with data collected from the interviews with higher education studies researchers.

It is acknowledged that data can be interpreted and classified along a continuum ranging from completely open to completely closed depending on the criteria used to assess openness and, ultimately, on how the authors of the assessment method understand how open data is expected to make an impact in a particular context. We have used both the 8 Principles of Open Government as well as the Exploring the Emerging Impacts of Open Data in Developing Countries project’s 10-point assessment framework in order to evaluate whether a data source is open or closed (ODDC 2013). Based on our experience in working with open data in South Africa, we are aware of extremely low levels of interoperability and general confusion in terms of open data licensing. We therefore determined that a data source had to score at least 80% on both of the open data assessment tools used for it to be considered open. In other words, if we included open licensing and interoperability in the scoring, and insisted that data sets scored 100% in order to be considered open, we would not have had a single open dataset in the ecosystem. We therefore felt it necessary to introduce a handicap system when evaluating the openness of data sources.

The sample was further refined by excluding data sources that (i) contained data on the public university sector but which were deemed to be irrelevant in the context of university governance, that is, they provided data that were unlikely to be used in university-level decision-making, planning and implementation; (ii) contained data only at the system or aggregated national level and not at the level of the university (or more granular than the university level).³² The justification for the exclusion of system-level data is that it is assumed that university planners, for purposes of the strategic planning of their own

³² It is acknowledged that datasets which provide a system-level view and are non-institution-specific may also be used in analysis and decision-making at institution level. Examples include data on population demographics, income and expenditure patterns, migration, etc. National data of this data resides with StatsSA. However, for the purposes of simplifying the analysis, the (open) data ecosystem was limited to data that provides at least institutional-level granularity.

institution, are less interested in data that describes the entire national public university system. As a fairly crude example: planners would rather know the enrolment numbers at their own and at other universities than know the total number of enrolments for all universities. In any case, the availability of institutional-level data allows for the calculation or extrapolation of the national picture by data users.

Policies and legislation

The identification of relevant policies and legislation was done by means of desk research. Ren's (2013) 'Opening public data in South Africa: Legal complications' provided a useful starting point in identifying relevant legislation. In our case, relevant policy and legislation was taken to mean those policies and laws which have a direct bearing on the provision and use of data in the governance of South African public universities, with a particular emphasis on any such policies or laws that have a bearing on open data.

Context

The ecosystem reflects three contextual conditions under which the actors function and which motivate, direct and/or constrain their actions as data providers, intermediaries or consumers.

The first of these is the regulatory condition ('R' in the diagram) – laws, policies, standards and agreements which have a bearing on how the components of the ecosystem are structured and how they interrelate. The ecosystem model indicates a plethora of acts as well as policies and other mandates which have a bearing on the actors in the ecosystem.

The second condition is that of the institutional context in which the actors operate. Institutional boundaries are indicated by circles in the diagram. The institutional context (universities are some of the oldest institutions in the world along with the church and the military) provide the taken-for-granted values, rules and norms shared by actors who operate within a particular institutional context (Scott 1995). These values, rules and norms inevitably propel and restrain the behaviours of actors in the ecosystem (Janssen et al. 2012).

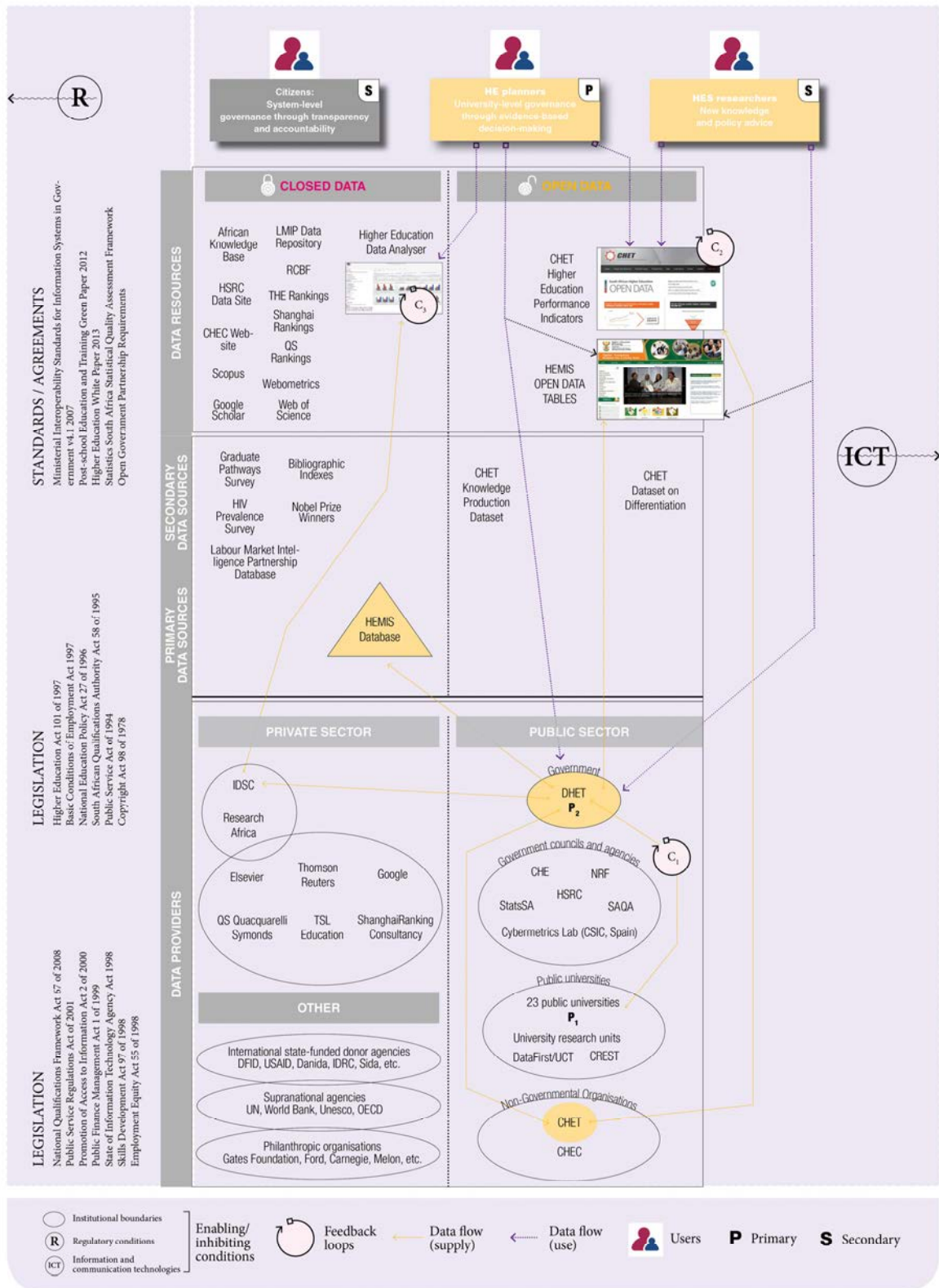
The third condition is that of current information and communications technologies, that is, the network elements, the network operators and the communications protocols that connect and interconnect the networked elements, operators and users ('ICT' in the diagram). Principle among these is the internet as a key enabler that introduces new actors to the ecosystem by connecting them to legacy components. Although in a strict sense a misnomer, the internet could very well be described as a keystone species in the open data ecosystem. Data could still flow without the internet, but its existence has facilitated the evolution of the ecosystem in ways not possible before.

The [open] data ecosystem

Our modelling of the wider public university governance data ecosystem, including, but not limited to, open data, is presented in Figure 4. It assumes an open system in which new actors operate in a previously closed, government-controlled information system. Traditional boundaries have been displaced and non-public actors have entered into the data system (Janssen et al. 2012).

Within the contextual forces, both enabling and restrictive, exerted by ICTs, institutions and the regulative environment, the ecosystem model locates the relative positions of the actors in the ecosystem: data providers, sources, resources and users. Whether actors are primary ('P') or secondary ('S') (based on Helbig *et al.* [2012]) and the flow of data between them is also represented. It is to this part of the ecosystem – what Fransman (2010:24) describes as the "economic-institutional" component of his ICT ecosystem layering model – that we pay most attention in our analysis.

Figure 4: The South African public university governance [open] data ecosystem



Also indicated in the ecosystem model are emerging feedback loops between supply and demand for open data in the ecosystem ('C₁', 'C₂' and 'C₃' in Figure 4): "A feedback loop exists when information resulting from some action within the system (endogenous) travels through the system and eventually returns in some form to its point of origin and potentially influences future action" (Helbig et al. 2012: 22).

Open data does not exist in isolation

Representing an ecosystem as a diagram tends towards a simplistic view, the inevitable result of analysis which attempts to create order out of a complex, non-linear set of processes, especially through the lens of a particular conceptual framework. Nevertheless, the representation is revealing: it shows that while there is a relatively even distribution of data providers in the ecosystem between public and private sectors, the number of data sources is unevenly distributed between those which are open and those which are closed. It is therefore not possible to describe an exclusively *open* data ecosystem.

Because there are many connections and interdependencies between closed and open data sources in the ecosystem (not all of which are necessarily captured in the simplified schema of the ecosystem model presented) one cannot capture the full range of dynamics at play in a data ecosystem that is only open. Also, as current research by Magalhaes *et al.* (2013) seem to suggest, innovators tend to draw on data from both closed and open resources in the development of applications. A by-product of this blended approach on the part of application developers is that it muddies the waters when trying to measure the impact of open data per se in innovation systems, on transparency and, ultimately, on social or economic development.

The preponderance of datasets remains closed using the Sebastopol Principles and the ODDC criteria, despite the healthy representation of the public sector on the supply side. A cautionary note: datasets indicated as closed are not necessarily inaccessible; they simply do not meet the criteria set out by the Sebastopol Principles and the ODDC. However, because they don't meet these criteria, the implication is that their impact is limited or stifled. Data that are not machine readable, interoperable, openly licensed, etc. – criteria which many of the datasets in our ecosystem do not meet – are more limited in their potential uptake and reuse. Planners (as the primary users in our ecosystem) draw on both open and closed datasets in the execution of their governance tasks, but their activities in accessing, collating and interpreting the data for the purposes of informed decision-making could be much more efficient and could yield new relational insights if more of the data was open, linked and licensed. The same applies to other data users in the ecosystem.

Helbig *et al.*'s (2012) information polity heuristic in which the primary source and resource are presumed to reside within government does not hold in our case. The primary data source (where data is collected and processed, indicated as 'P₁') is located outside of government (in the universities) while the corresponding primary data resource (HEMIS) is hosted and maintained by government. We indicate government as a second-level primary data source in the ecosystem ('P₂'). In other words, when extending the field of analysis from government to the public sector, the location of primary data sources could be in autonomous public bodies (such as universities). It is therefore suggested that analyses of open data ecosystems in relation to governance not be conflated with government – additional governance domains are likely to exist in the broader ecosystem and may have a bearing on how government open data is supplied and (re)used.

Power dynamics

What Helbig's (2012) information polity retains, and which we would suggest the ecosystem presented here lacks, are the power dynamics at play between actors. The marketisation of higher education and the fears of primary data providers related to how their data will be used has been alluded to elsewhere in

this report. This points to the relatively powerful position of primary data providers in the ecosystem. The ecosystem does not capture the power relations between primary providers and other actors in the ecosystem, nor does it reflect the extent to which citizens are able to mobilise in order to counter the power of primary providers of data. If injustice is seen as a potential outcome of an imbalance in power, then Johnson (2013) is correct to caution about the possible injustices inherent in open data provision. Johnson (2013), approaching the use of open data from the demand-side, expresses concerns over the correlation between open data and information injustice, a scenario premised on the differential capabilities of users; capabilities which result in an uneven distribution of power between users. Again, the ecosystem does not capture the power dynamics at play between users – a condition which may well determine the (re)use and, more importantly, the impact of open data.

Keystone species

“Keystone species” are considered crucial to ecological functioning because their presence performs some vital enabling function (Nardi and O’Day 1999: 53), either as mediators, actors who bridge institutional boundaries and translate across disciplines, or by creating value for their ecosystems by creating platforms, services, tools or technologies that offer solutions to others in the ecosystem (Iansiti and Levin 2004: 7).

CHET enables new connections and solutions within the ecosystem. For example, while university planners can access the anonymised HEMIS data tables from the DHET website, the CHET open data platform enables planners to compare themselves with other universities across a set of indicators using the tools developed for doing so. CHET is also located outside of the two primary institutions – the state and the university – thus enabling it to play a mediating role. CHET as intermediary therefore plays a vital role in the ecosystem, and could be described as a keystone species within the South African public university open data ecosystem.

Keystone species are enablers, not necessarily drivers in the ecosystem; they can be useful but they are not essential to the sustained functioning of an ecosystem. The public university system is a competitive landscape in which all 23 public universities compete for finite resources (such as fee-paying students, government block grants, research project funding, etc.). In this context, new knowledge has value in that it may inform decisions and implementation to give a university a competitive advantage over its rivals. Innovation can therefore be seen as the key driver in the ecosystem as there is virtuous circle between data production, data supply (open or closed) and consumption.

Sustainability

Whether this virtuous cycle is sustainable in the case of open data supply and consumption is less clear. The collection, repackaging and provision of data in a format and context that ensures greater probability of use and impact (Helbig *et al.* 2012) requires the investment of resources (Iansiti and Levin 2004). External funding, predominantly from foreign philanthropies, ensures the ability of the intermediary (CHET) to provide *open* data. The feedback loop (‘C₂’) is reinforced if there is evidence of use/impact as this increases the likelihood of future external funding but does not guarantee it. The provision and impact of open data by the intermediary in the ecosystem is therefore not inherently sustainable.

The issue of sustainability is highly relevant given the presence of a second intermediary in the ecosystem. As O’Neil (2013: 33) states unequivocally: “Without money, there is no sustainability.” IDSC is commercial supplier of public university governance data via its Higher Education Data Analyzer (HEDA) platform. South African universities pay annual subscription fees to access IDSC’s HEDA data. IDSC’s presence in the ecosystem suggests that in the case of public university governance in South Africa, data users (in their case, universities) do, in fact, derive value from the data and that they are prepared to

enter into an exchange relationship for the provision of data. Similar business models exist in other countries (e.g. Academic Analytics in the US). For as long as the users perceive value in the provision of data, and are adequately resourced to enter into an exchange relationship,³³ this part of the ecosystem appears to be more stable and sustainable ('C₃').

An unfortunate irony would be if the less-sustainable actor – CHET – disrupted the closed data system only for the vacuum in the ecosystem to be filled by a more financially sustainable actor – IDSC. Stated differently, a less sustainable virtuous cycle premised on openness could potentially initiate a more sustainable vicious cycle in which public data is used for private gain. This raises serious questions about the viability of not-for-profit, civic-minded intermediaries and, ultimately, the sustainability of *open* data ecosystems; particularly those ecosystems in which the state plays a weak supportive and co-ordinating role in the ecosystem.

For the time-being both intermediaries co-exist in the ecosystem. Given that one of the intermediaries is less sustainable, one future scenario in this particular ecosystem would be the disappearance of this intermediary, the continued presence of the commercial intermediary and a concomitant closing of the data in the system. A second future scenario in other ecosystems where intermediaries are civic start-ups rather than non-governmental organisations, is the possibility of these start-ups being bought out by larger commercial enterprises as they detect value in the underlying user data collected (rather than in the social service developed). In a developing country context where markets are both large and untapped, the sustainability of intermediaries is a reason for concern if the value of open data is to be realised. Philanthropic support may be providing the impetus, but capital will have to flow, and governments will have to the conditions for both the economic and social benefits of open data to be realised.

Capacity

It is to the role of the government as a provider of data in the ecosystem that attention is now turned. Government, in the form of DHET, is a third data provider in the data supply chain. As in the case of CHET, DHET provides open data. In the case of DHET, the provision of open data is not driven by financial incentives or rewards, nor by civic-mindedness. Based on the interviews conducted, the provision of open data appears to be driven by a need for efficiency due to limited capacity in the provision of open data to a broad user base. It is critical to note the implications of identifying the real driver for data provision at government level – existing policies (such as interoperability standards), agreements (such the Open Government Partnership) or legislation, of which there is a plethora, have little bearing on the provision of open data. The practical, day-to-day realities of those tasked with managing the data, appears to trump the dictates of policy and even legislation.

Furthermore, while government has opened up the HEMIS data by providing the data tables on its website, the impact of this action in creating new connections between the open data and potential users is minimal. Interviews confirm that university planners, industry, supranational agencies, higher education studies and other researchers, the media and other stakeholders still approach DHET directly for data, rather than downloading the data from the DHET website. The provision of open HEMIS data appears to have had little impact in disrupting this behaviour on the part of data consumers in the ecosystem. In terms of Ding et al.'s (2012) roadmap of linked open data, DHET has entered the 'open stage', and still has some way to go before proceeding to the 'link stage'. A shift from open to linked data may be a necessary step in improving levels of open data use. DHET's as both the primary data source (HEMIS) as

33 It is worth noting from our findings on the use of higher education data, that higher education researchers do not make use of HEDA. Researchers, some of which are independent or employed by non-governmental organisations, most likely do not have the financial resources to subscribe to the HEDA platform.

well as an open data resource (online HEMIS data tables), places it in a unique position in the ecosystem to facilitate access to and to unlock the full network benefits of open data.

However, a first, more fundamental step may be needed to increase the use of DHET's open data. Ding *et al.*'s (2012) open stage stipulates not only that governments place datasets online but that they assist citizens in finding relevant datasets. And Helbig *et al.* (2012) in their case study research, highlight the importance of the context in which data is shared as a determinant in the uptake of open data by consumers. Context is a determinant factor in avoiding conflicts of meaning, misinterpretation and user frustration. Unless data providers (including government) not only pay attention to but invest resources to create contexts in which open data is easy to interpret and consume, open data initiatives risk reducing their impact on governance as well as their contribution to innovation and socio-economic development.

Huijboom and Van den Broek (2011) in their five-country case study on open government data strategies, identified ten drivers and ten barriers to open data policy implementation. All five countries cited inspiring strategies of other countries as an important driver. When interviewed, DHET referred to the US's Integrated Postsecondary Education Data System (IPEDS)³⁴ as the ideal platform and as something for the Department to aspire to. In terms of barriers to opening government data, our case supports Huijboom and Van den Broek's (2011) finding that barriers tend to be located within government itself and the findings of Janssen *et al.* (2012) that such barriers are institutionally bound. In our case, we identified six such barriers: a closed government culture, privacy legislation, limited user-friendliness, lack of standardisation of open data policy and existing charging models. However, what we found to be the most restrictive barrier to open data implementation by government – capacity – was not mentioned by any of the developed countries in Huijboom and Van den Broek's (2011) study.

Returning to DHET's unique position in the ecosystem as the primary data source, it is noticeable from the ecosystem diagram that the public HEMIS database (represented by a triangle in Figure 4) is isolated from all other data resources in the ecosystem. Data from the HEMIS database shared with secondary data providers such as IDSC and CHET, are supplied by the four staff at DHET who have access to HEMIS. No external database or system draw data from HEMIS.³⁵ This seems surprising giving government's political commitment to e-government. It seems that while there is intent on the part of government, until government open data is supplied in an information context that meets the needs of its citizenry (constituted of a range of user types, both in terms of needs and levels of access) and is made available via platforms that allow for interoperability, the reuse of government with the open data ecosystem will remain limited. As Gurstein (2011) cautions: "Any critical analysis of 'open data' use has to include how and under what conditions the data that is being made available is contextualized and given meaning."³⁶ The limitations introduced into the ecosystem by how government currently supplies its data, is supported by the findings of Sharif (2013) in his research on how two South African research organisations access public sector information. Critically, as Sharif (2013) points out, the data collected and provided by government, despite being near-impossible to access or use, is nevertheless regarded

34 <http://nces.ed.gov/ipeds/datacenter/>

35 While it is acknowledged that the HEMIS database contains the personal information of university students and staff, it is presumed that using permissions and access controls, the bulk of the HEMIS data could be shared without compromising the personal data stored in the database.

36 Here we are only referring to the context in which the data is provided (a website, an online platform, a dashboard, etc. and the content they contain). Equally important, according to Gurstein (2011) are the variable contexts in which the spectrum of open data consumer finds themselves. A point echoed by the political commentator Steven Friedman when commenting on the implementation failures of the South African government's Open Government Partnership action plan – "Democratic government is meant to serve the people. This possibility is restricted when government alone decides the forums in which citizens should talk to it" (Friedman 2013).

as unique and valuable by the participants in his study. The HEMIS data is similarly unique and valuable to the constituent of users in the public university governance ecosystem.

Whether it is providing a richer information context to ensure greater interpretability or creating a primary data source that is interoperable, intermediaries have a valuable contribution to make in providing capacity and flexibility to resource and institutionally constrained government departments such as DHET.

Open government data initiatives are often linked to the notion of 'government as a platform' (O'Reilly, (2010) in which government acts as the primary data provider and innovative actors external to the state (re)use open government data to provide better, more efficient or more customised public services. In the case of this research project, these actors are the two intermediaries in the ecosystem and their presence attests to their greater (innovative) capacity over the public sector. However, to realise and to maximise fully the contribution of these intermediaries in the evolution of the ecosystem, government needs to interact with the intermediaries in the ecosystem. In our case, it is clear that there is little interaction between DHET and the two intermediaries in the ecosystem, and that this is inhibiting the evolutionary pace of the ecosystem.

Intermediaries and information injustice

In the ecosystem presented here, funding incentives to universities ensure the provision of data from universities to government. In other words, an incentive is already built into the ecosystem to ensure data capturing at institutional level and its supply to a central point located in government (i.e. DHET) (C₁); a condition which is absent in many other African data governance ecosystems (Cloete *et al.* 2011), and may be absent in other developing-country contexts. Again, a lack of capacity could be the primary constraint on open data provision in these contexts, and a lack of incentives or rewards is likely to maintain the status quo. However, incentives and rewards may introduce unintentional bias in the data collected.

DHET indicated that a process was underway to ensure that the Department's data complied with Statistics South Africa's South African Statistical Quality Assessment Framework (SASQAF). Compliance with SASQAF would mean that the DHET's higher education data would be elevated to official, national data status. Such compliance with SASQAF as well as the incentives present in the ecosystem for universities to capture data according to HEMIS specification, highlights the potential danger of reinforcing social injustices predicated on what Johnson (2013:12) refers to as 'disciplinary power' in the ecosystem:

[t]he opening of data can function as a tool of disciplinary power. Open data enhances the capacity of disciplinary systems to observe and evaluate institutions' and individuals' conformity to norms that become the core values and assumptions of the institutional system whether or not they reflect the circumstances of those institutions and individuals. Both individuals who deviate from these norms and the institutions that specialize in serving them are marginalized in policy debates; the surveillers and sousveilleurs evaluate all institutions according to the norm (and indeed data may only exist regarding it), and the institutions internalize the norms and orient their actions to them. With the norms reflecting the power structure of the society in which they developed, they reiterate the injustices that open data set out to ameliorate.

Implicit in understanding the functioning of open data in society is a sensitivity to institutions as sites of shared norms and values where conformity is prized. The open data movement needs to take heed of the institutional context when evaluating the effects and impacts of opening up data. A more nuanced

appreciation of institutional contexts will allow the open data movement to predict with greater certainty the possible strategic responses of those institutions (such as government) being pressured to open their datasets.

Johnson (2013) refers to the US IPEDS in imposing institutional conformity through disciplinary surveillance; HEMIS has a similar impact on South African public universities. This condition seems to highlight the importance of intermediaries in curtailing the 'de-ameliorating' effects of disciplinary surveillance on open data. In other words, intermediaries, as actors who may well operate outside of the boundaries of the state apparatus and of the institution of the university, have the propensity to challenge how data is collected, interpreted and shared. Their role as de-institutionalised actors could go some way in restoring the democratic value of open data. In addition, intermediaries are in a position to add to existing datasets thereby extending both the corpus of open data on higher education and the possibilities for new interpretations of the data. In our case, there is evidence of this expansion in two forms. First, CHET's open data platform provides financial data which is neither collected by HEMIS nor shared by DHET. Second, because CHET provides indicator data, it has introduced new forms of analysis (e.g. on success rates or cost per graduate) which challenge the normative assumptions inherent in the DHET's construction of the HEMIS database.

It goes without saying that CHET, in the process of representing the HEMIS data and by adding its own data, is not immune to embedding its own values into the open data presented. However, as Johnson (2013) suggests, pluralism is one approach to countering information injustices. By promoting multiple, even conflicting, information systems, by including multiple stakeholders in the design of such systems and by broadening the range of data analysers, the undesirable effects of embedded norms and values are more likely to be ameliorated. Intermediaries, it would appear, have an important role to play in this regard.

Conclusion

This section of the report set out to establish whether government-supplied open data is viscous or fluid, whether its data is contributing to the evolution of the ecosystem, and what role intermediaries are playing, if any, in the evolution of the South African higher education governance open data ecosystem.

From the interviews conducted, the South African government (as represented by DHET) makes open data available because of capacity constraints. Rather than imperatives of transparency or accountability, it is an efficiency imperative which has opened up HEMIS data. But uptake of the HEMIS open data on the DHET website appears to be weak; an array of data users still approach DHET directly for data from HEMIS.

In parallel to the DHET's provision of open data, data is provided by two intermediaries in the ecosystem – CHET via the open data tool on its website and IDSC via online, subscription-based HEDA dashboards. These intermediaries follow different modes of data provision – one open and one closed. Both have a very clearly defined target audience in the form of university planners and both provide relatively elaborate information contexts with these users in mind. In contrast, it would be difficult to describe the DHET open data information context, as it currently exists on its website, as user-centric. Returning to our liquid analogy, the DHET open data is viscous rather than fluid.

The net result of the viscous DHET open data is that while DHET opens up HEMIS data to relieve capacity constraints, opening up the data has not alleviated the demands made on the limited number of people in the department who have access to the data. While it is true that the ecosystem has evolved due to the activities of the intermediaries and in spite of the DHET open data's poor information context, DHET would do better (i) by improving the information context in order to facilitate the uptake and (re)use

of its open data; (ii) by making the HEMIS data interoperable, either directly or indirectly, thereby allowing more interest and activity from existing and from new intermediaries; and (iii) by engaging with intermediaries in the ecosystem in order to make the most of their innovation, capacity and flexibility. Such moves will contribute to the evolution of the public university governance ecosystem by decreasing the viscosity of government-supplied open data and increasing the fluidity of open data between actors in the ecosystem. This will sustain and promote plurality in the supply of open data and increase responsible use by university planners, higher education studies researchers and other stakeholders in their efforts to ensure good governance in South Africa's public universities.

Intermediaries, driven by civic or financial imperatives, and using their social capital to access data directly from government, are nevertheless creating their own platforms to make valued data (open or closed) pertinent to the governance of South Africa public universities available to clearly identified user groups within the public university governance ecosystem. It is in this part of the ecosystem that we see expansion; expansion that could be propelled by government. Time will tell if these new platforms are effective, efficacious and sustainable.

Section 7. Summary of findings

1. CHET's open data is being used by university planners and by higher education studies researchers, albeit infrequently. Researchers expressed the need for richer, more granular data. Both planners and researchers found value in CHET's open data platform allowing them to make comparisons between universities. CHET would do well to take note of these findings when planning any modifications to its open data platform. In particular, it may need to carve a more differentiated role based the needs of different user groups and given the presence of other more financially secure intermediaries in the open data ecosystem.
2. HEMIS is an isolated data source. Granting access to HEMIS by third-parties (under controlled conditions to protect personal data) could further stimulate the evolution of the open data ecosystem and relieve pressure on capacity-constrained government departments. This could strengthen the impact of open data on the governance of South African public universities. Discussions between DHET and other stakeholders, including the current intermediaries in the ecosystem, on how to share HEMIS data and how to improve the interpretability of the open data tables currently made available online, could stimulate a new phase in the evolution of the ecosystem.
3. There are concerns at both government and university levels about how data will be used and (mis)interpreted, and this may constrain future data supply. Education both at the level of supply (DHET) and at the level of use by the media in particular on how to improve the interpretability of data could go some way in countering current levels of mistrust. Similar initiatives may be necessary to address uneven levels of data use and trust apparent across university executives and councils.
4. Open data intermediaries increase the accessibility and utility of data. While there is a rich publicly-funded dataset on South African higher education, the data remains largely inaccessible and unusable to universities and researchers in higher education studies. Despite these constraints, the findings show that intermediaries in the ecosystem are playing a valuable role in making the data both available and useable.
5. Open data intermediaries provide both supply-side as well as demand-side value. CHET's work on higher education performance indicators was intended not only to contribute to government's steering mechanisms, but also to contribute to the governance capacity of South African universities. The findings support the use of CHET's open data to build capacity within universities. Further research is required to confirm the use of CHET data in state-steering of the South African higher education system, although there is some evidence of CHET's data being referenced in national policy documents.
6. Intermediaries may assume the role of a 'keystone species' in a data ecosystem.
7. The findings show that intermediaries such as CHET play an enabling role of mediation and innovation within the ecosystem. CHET enables new connections and solutions within the ecosystem. CHET is also located outside of the two primary institutions – the state and the university – thus enabling it to play a mediating role.
8. Intermediaries democratise the effects and use of open data. Intermediaries play an important role in curtailing the 'de-ameliorating' effects of disciplinary surveillance on open data.

Intermediaries, as actors who may well operate outside of the boundaries of the state apparatus and of the institution of the university, have the propensity to challenge how data is collected, interpreted and shared. Their role as de-institutionalised actors could contribute to restoring the democratic value of open data. The findings show that CHET is already playing a unique role in ensuring information justice as it challenges existing, imposed norms in the collection and use of open data in the governance of South Africa's public university system.

The South African higher education governance open data ecosystem has evolved despite poor data provision by government because of the presence of intermediaries in the ecosystem. By providing a richer information context and/or by making the data interoperable, government could improve the uptake of data by new users and intermediaries, as well as by the existing intermediaries. Increasing the fluidity of government open data could remove uncertainties around both the degree of access provided by intermediaries and the financial sustainability of the open platforms provided by intermediaries.

References


- Amaral A, Jones GA & Karseth B (2002) Governing higher education: National perspectives on institutional governance. In Amaral A, Jones GA & Karseth B (eds), *Governing Higher Education: National Perspectives on Institutional Governance*. Dordrecht: Kluwer. pp. 279–298.
- Bauer F & Kaltenböck (n.d.) *Linked open data: The essentials*. Vienna, Edition mono/monochrom.
- Bunting I (2010) *Performance Indicators: The South African Higher Education System 2000-2008*. Cape Town: CHET.
- Bunting I & Cloete N (2004) *Developing Performance Indicators in Higher Education*. Cape Town: CHET
- Bunting I, Sheppard C, Cloete N & Belding L (2010) *Performance Indicators: South African higher education 2000-2008*. Cape Town, CHET.
- Bentley K, Habib A & Morrow S (2006) *Academic Freedom, Institutional Autonomy and the Corporatised University in Contemporary South Africa*. Pretoria: Council on Higher Education.
- Brennan J (2008) Higher education and social change. *Higher Education* 56(3): 381–393.
- Castells M (2001) Universities as dynamic systems of contradictory functions. In Muller J, Cloete N & Badat S (eds), *Challenges of Globalisation: South African Debates with Manuel Castells*. Cape Town: Maskew Miller Longman.
- Clark, BR (2008) The Advantages of Case Study Narratives in Understanding Continuity and Change in Universities. In B.R. Clark, *On Higher Education: Selected Writings, 1956-2006*. Baltimore, Johns Hopkins University Press. 549-554.
- Cloete N & Bunting I (2000) *Higher Education Transformation: Assessing Performance*. Cape Town: CHET.
- Cloete N, Bailey T, Pillay P, Bunting I & Maasen P (2011) *Universities and economic development in Africa*. Cape Town: Centre for Higher Education Transformation (CHET).
- Datacatalogs.org. Accessed 23 July 2013. <http://datacatalogs.org/>
- Datahub. Accessed 23 July 2013. <http://datahub.io/dataset>
- Davies T (2013) Assessing open data supply: Challenging questions. 8 May. Accessed 23 July 2013. <http://www.opendataresearch.org/news/2013/assessing-open-data-supply-challenging-questions>
- Davies T, Perini F & Alonso JM (2013) *Researching the Emerging Impacts of Open Data: ODDC Conceptual Framework*. July 2013 – ODDC Working Papers #1.
- Department of Higher Education and Training (Republic of South Africa) (2013) White Paper for Post-school Education and Training: Building an Expanded, Effective and Integrated Post-school System. Pretoria: Department of Higher Education and Training.
- Ding L, Peristeras V & Hausenblas M (2012) Linked Open Government. *IEEE Intelligent systems* May/June 2012: 11-15.
- Exploring the Emerging Impacts of Open Data in Developing Countries (ODDC). Draft framework: Assessing country level open data supply: framework and research strategies. Accessed 23 July 2013. https://docs.google.com/document/d/1z-T3QmmZTmWkFrKySi-x_EBDVqlt6Orr6z-C58gPGeg/edit
- Fransman M (2010) *The New ICT ecosystem: Implications for policy and regulation*. Cambridge: Cambridge University Press.
- Gurstein M (2011) Open Data: Empowering the empowered or effective data use for everyone? *First Monday* 16:2.

- Harrison C, Pardo TA & Cook M (2012) Creating open government ecosystems: A research and development agenda. *Future Internet* 4:900-928. Doi:10.3390/fi4040900
- Higher Education Data Analyzer (HEDA). Accessed 25 July 2013.
<http://www.heda.co.za/pds/members/background.aspx>
- Helbig N, Cresswell AM, Burke GB & Luna-reyes L (2012) *The Dynamics of Opening Government Data*. New York, Center for Technology in Government.
- Huijboom N & Van den Broek T (2011) Open Data: An International Comparison of Strategies, *European Journal of ePractice* 12: 1-13.
- Iansiti, M. & Levin, R. (2004). Strategy as ecology. *Harvard Business Review*, 1 March 2004, pp. 68–78.
- Intarakumnerd P & Chaoroenporn P (2013) The roles of intermediaries in sectoral innovation system in developing countries: public organizations versus private organizations. *Asian Journal of Technology Innovation*, 21:1, 108-119.
- International Development Research Centre Act. R.S., c. 21(1st Supp.), s. 1. Accessed 24 October 2013.
<http://laws-lois.justice.gc.ca/PDF/I-19.pdf>
- Janssen M, Charalabidis Y & Zuiderwijk A (2012) Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management* 29: 258–268.
- Johnson JA (2013) From Open Data to Information Justice. Paper presented at the Annual Conference of the Midwest Political Science Association, 13 April 2013, Chicago.
- Linked Open Data (LOD) Cloud. Accessed 23 July 2013. <http://www.lod-cloud.net/state/>
- Magalhaes G, Roseira C & Strover S (2013) Open government data intermediaries: A terminology framework. In Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance, ICEGOV2013, Seoul, Republic of Korea, 22-25 October 2013. ACM Press.
- Manyika J, Chui M, Groves P, Farrell D, Van Kuiken S & Doshi EA (2013) Open data: Unlocking innovation and performance with liquid information. McKinsey Global Institute. Accessed 1 November 2013.
http://www.mckinsey.com/~media/McKinsey/dotcom/Insights/Business%20Technology/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_OpenData_Full_report_Oct2013.ashx
- Matheka R (2013) Independent Reporting Mechanism: South Africa progress report 2011–2013. Open Government Partnership. Accessed 16 October 2013.
<http://www.opengovpartnership.org/country/south-africa>
- McAuley D, Rahemtulla H, Goulding J & Souch C (2011) How open data, data literacy and linked data will revolutionise higher education. In: Coiffait L (ed.) *Blue Skies: New Thinking about the Future of Higher Education*. London: Pearson Centre for Policy and Learning. pp. 88–93.
http://pearsonblueskies.com/wp-content/uploads/2011/05/21-pp_088-093.pdf
- Muller J & Cloete N (1986) The white hands: Academic social scientists and forms of popular knowledge production. *Critical Arts* 4(2): 1-19.
- Nardi, B. & O'Day, V.L. (1999). *Information ecologies: Using technology with heart*. Cambridge, MA: MIT Press.
- Ncayiyana DJ & Hayward FM (1999) *Effective governance: A guide for council members of universities and technikons*. Cape Town, CHET.
- O'Neil DX (2013) Building a Smarter Chicago. In: Goldstein B & Dyson L, *Beyond Transparency: Open Data and the Future of Civic Innovation*. San Francisco, Code for America Press.
- O'Reilly T (2010) Government as a platform. In D Lathrop & L Ruma (eds) *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

- Open Definition. Accessed 23 July 2013. <http://opendefinition.org/>
- Open Government Data. Accessed 23 July 2013. <http://www.opengovdata.org/home/8principles>
- Pillay P (2010) South Africa. In: Pundy, P. (ed.) *Higher education financing in East and Southern Africa*. Cape Town, African Minds.
- Reed MI, Meek VL & Jones GA (2002) Introduction. In Amaral A, Jones GA & Karseth B (eds), *Governing Higher Education: National perspectives on institutional governance*. Kluwer. pp. xv–xxxi.
- Rens A (2013, April 15) Opening public data in South Africa: Legal complications. Accessed 2 November 2013. <http://aliquidnovi.org/opening-public-data-in-south-africa-legal-complications/>
- Republic of South Africa (2011) Open Government Partnership Action Plan: Republic of South Africa. Accessed 2 November 2013. <http://www.opengovpartnership.org/country/south-africa/action-plan>
- Scott WR (1995) *Institutions and organizations*. Thousand Oaks, CA: Sage.
- Statistics South Africa (2010) South African Statistical Quality Assessment Framework (SASQAF). 2nd Edition. http://beta2.statssa.gov.za/standardisation/SASQAF_Edition_2.pdf
- Sharif RM (2013) Utilization and Value of Public Sector Information for Knowledge Development: the Case of South Africa. Unpublished PhD thesis, Syracuse University.
- Stensaker B (2002) Trance, Transparency and Transformation: The impact of external quality monitoring on higher education. Keynote presentation at the Seventh Quality in Higher Education International Seminar, Melbourne, 29–31 October 2002
- Stensaker B (2003) Trance, Transparency and Transformation: The impact of external quality monitoring on higher education. *Quality in Higher Education* 9(2): 151–159. *Sunday Independent* (2012) 'Poor leadership cripples tertiary institutions', 29 July.
- UN ESCAP (n.d.) What is Good Governance? <http://www.unescap.org/pdd/prs/ProjectActivities/Ongoing/gg/governance.asp>
- Van Schalkwyk F (2013) Supply-side variants in the supply of open data in South African public university governance. In T Janowski, J Holm & E Estevez (Eds), *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*, Seoul, Republic of Korea, 22–25 October 2013 (ICEGOV2013). ACM Press, International Conference Proceedings Series.
- World Wide Web Foundation (2013) Web for civic engagement. Accessed 24 October 2013. <http://www.webfoundation.org/initiatives/web-for-civic-engagement/>
- Yu H & Robinson DG (2012) The New Ambiguity of "Open Government", *UCLA Law Review Discourse* 178: 178–208.

Appendix 1

The CHET Open Data initiative (<http://www.chet.org.za/data>)



Web update on Open Data

Tweet

CHET'S ONLINE OPEN DATA SETS

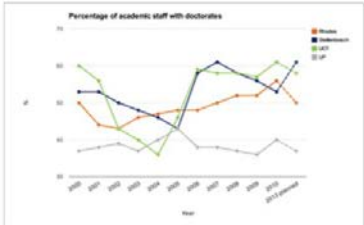
CHET has made 3 data sets available on its website:

1. Performance indicator data on the 23 universities comprising the South African higher education system;
2. Data by province on South Africa's further education and training (FET) sector; and
3. Performance indicator data on eight African universities from eight African countries.

>> www.chet.org.za/data

South African Higher Education Data


Data updated to include 2010 HEMIS data. Data includes DHET-approved 2013 targets. Graphs redesigned for improved readability. All graph-related data downloadable in Excel.



African Higher Education Data

The data presented here is drawn from the most recent publication on the subject, *Cross-National Performance Indicators: A case study of eight African universities*.

The data presents for the first time a set of cross-national, comparative data on the performance of eight African universities.



South African FET Colleges Data

The data on the South African post-secondary education sector presented is drawn from the CHET publication, *Shaping the Future of South Africa's Youth: Rethinking post-school education and skills training*.

The data is made available via Google's Public Data service which affords users the ability to create fully customised and comparative graphs using a rich set of variables.

▼ South African Further Education...

- ▶ 01 No. of colleges and camp...
- ▶ 02 Enrolments (public vs priv...
- No. of students enrolle...
- ▼ 03 Qualifications distribution...
- N certificate level grou...
- ▶ 04 Student age distribution
- ▶ 05 Staff composition by job t...
- ▶ 06 Staff distribution by job ty...
- ▶ 07 Staff remuneration by job ...
- ▶ 08 Qualification awarded

Clear

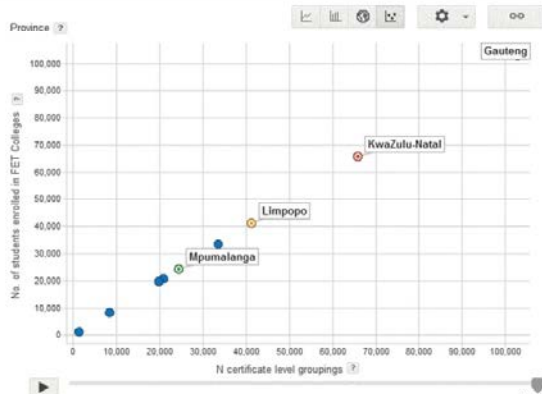
▼ Compare by Province ▼

- ☐ Eastern Cape
- ☐ Foreign
- ☐ Free State
- ☒ Gauteng
- ☒ KwaZulu-Natal
- ☒ Limpopo
- ☒ Mpumalanga
- ☐ North West
- ☐ Northern Cape
- ☐ Western Cape

Clear selections

▼ Public / Private

▼ Qualifications - Grouped



Data from Centre for Higher Education Transformation - Powered by Google


2009 [Explore data](#)

Having trouble reading this? [View it in your browser](#). Not interested anymore? [Unsubscribe instantly](#).

South African higher education performance data 2000 to 2010

In order to generate a graph, follow 5 simple steps:

1. Click on "Click here to create your graph" below and select an indicator from the drop-down list which appears.
2. Narrow the indicator variables (if applicable).
3. Select up to 4 universities.
4. Generate the graph (and relevant explanatory notes/definitions).
5. View or download the graph, download the data in an Excel spreadsheet, or comment on the graph.

 [Click here to create your graph](#)

About CHET's on-line higher education performance indicators

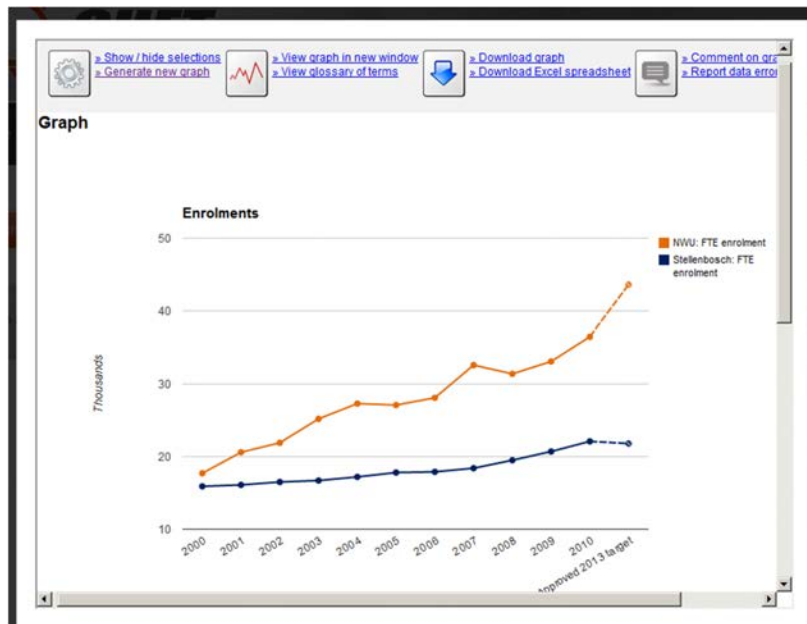
The 1997 White Paper on Higher Education Transformation and the 2001 National Plan for Higher Education indicated that the South African government would develop steering mechanisms, involving planning and funding, to assist with the transformation of the public higher education system. For this to function effectively, it is crucial that university councils understand how performance will be measured.

CHET's experience has been that this will not be an easy task for many university councils; currently very few institutions produce data sets which allow council members to engage meaningfully in discussions about the performance of their institution.

CHET has therefore decided to make performance indicator data available on-line to enable university councils to better assess their performance against the Minister's targets, to their own institutional targets and to the performance of their peers. This data tool will also be of use to higher education analysts, policy-makers and other decision-makers seeking a more detailed, empirically-based picture of South African higher education.

Disclaimer

Data for CHET's online performance indicators were obtained from the Higher Education Management Information System (HEMIS) of the Department of Higher Education and Training. The quality of the HEMIS data is dependent on the accuracy of the institutional data submissions.



Appendix 2

Selection matrix for university planning units to include in the sample

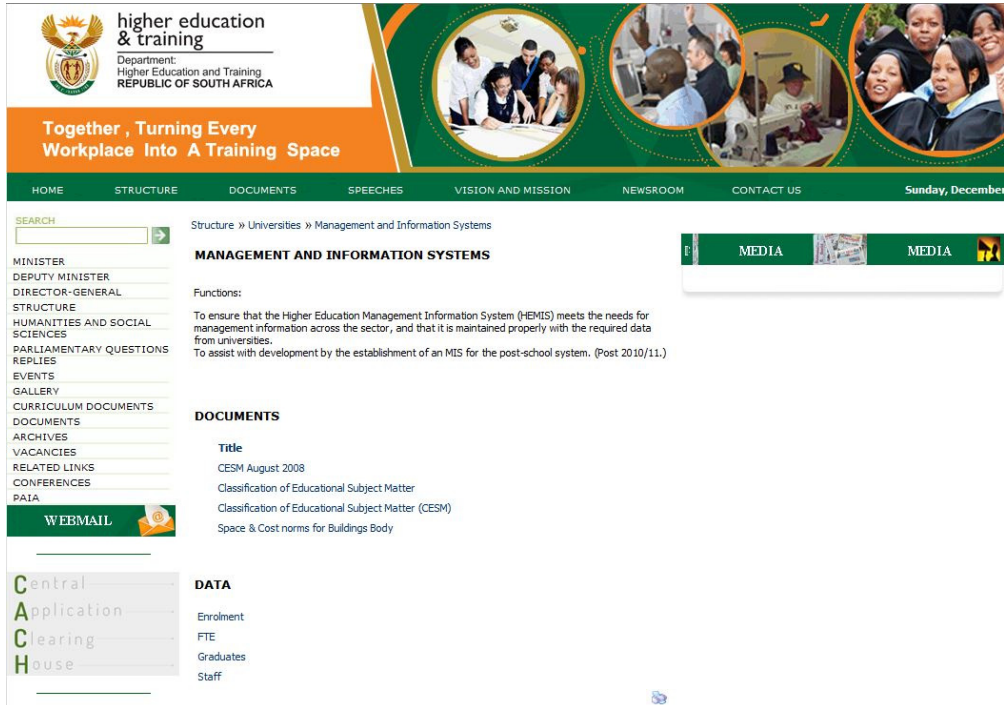
University	Size			Type				Location		Campus		Governance crisis
	Small	Medium	Large	Research	UoT	Comp.	Distance	Urban	Rural	Single	Multi	
CPUT		X			X			X			X	
CUT	X				X				X			X
DUT		X			X			X			X	
MUT	X				X				X	X		
NMMU		X				X		X			X	
NWU			X	X				X	X		X	
Rhodes	X			X					X	X		
Stellenbosch		X		X				X		X		
TUT			X		X			X	X		X	X*
UCT		X		X				X		X		
Fort Hare	X			X					X		X	
UFS		X		X				X	X		X	
UJ			X			X		X		X		
UKZN			X	X				X			X	
UL		X		X					X			
UP			X	X				X		X		
Unisa			X				X					
UV	X					X			X			
UWC	X			X				X		X		
Wits		X		X				X		X		
UZ	X					X			X			X
VUT		X			X			X				X
WSU		X				X		X			X	X
NOTES	Full-time equivalent enrolments <15,000 15-30,000 >30,000			As defined by the DHET UoT = University of Technology Comp. = Comprehensive university						Multi = 2nd campus more than 100km from central admin campus		Placed under administration by Treasury

7

Selected universities

Appendix 3

The DHET open data website



The screenshot displays the DHET open data website. The header features the DHET logo and the tagline "Together, Turning Every Workplace Into A Training Space". The navigation menu includes links to HOME, STRUCTURE, DOCUMENTS, SPEECHES, VISION AND MISSION, NEWSROOM, and CONTACT US. The main content area is titled "MANAGEMENT AND INFORMATION SYSTEMS" and includes a search bar, a list of links (MINISTER, DEPUTY MINISTER, DIRECTOR-GENERAL, etc.), a "WEBMAIL" button, and a "Central Application Clearing House" button. The "DOCUMENTS" section lists titles such as "CESM August 2008" and "Classification of Educational Subject Matter (CESM)". The "DATA" section lists categories like "Enrolment", "FTE", "Graduates", and "Staff".



The screenshot displays the DHET open data website, specifically the "ENROLMENT" section. The header and navigation menu are identical to the previous screenshot. The main content area is titled "ENROLMENT" and includes a search bar, a list of links, a "WEBMAIL" button, and a "Central Application Clearing House" button. The "ENROLMENT" section lists titles such as "2012 Table 2.12 for all institution" and "2012 Table 2.7 for all institution".